

理与人

〔英〕德里克·帕菲特 著　王新生 译　黄颂杰 校

二十世纪西方哲学经典

Derek Parfit

Reasons and Persons

上海译文出版社

版权信息

标题：理与人

作者：〔英〕德里克·帕菲特

译者：王新生

关注微博：@数字译文

微信订阅号：数字译文

联系我们：hi@shtph.com

问题反馈：complain@shtph.com

合作电话：021-53594508



Digital Lab是上海译文出版社数字业务的实验部门，成立于2014年3月。我们致力于将优质的资源送到读者手中。我们会不断努力，做体验更好，设计更好的电子书，加油！

上海译文出版社 | Digital Lab

中译本前言

《理与人》是由英国当代道德哲学家、心灵哲学家、牛津大学万灵学院研究员德里克·帕菲特撰写的一部道德哲学、伦理学著作，由牛津大学出版社出版，1984年首版，1986年、1987年相继出版修订本。这里我们就为什么要译这本书以及这本书的主要内容、观点作简要的说明和评述。

一

我在从事西方哲学的教学和科研工作中，十分关注基督教和伦理学、道德哲学，而且总喜欢把基督教和伦理道德联系起来。在我看来，宗教和道德是理解和把握西方哲学乃至西方文化的关键。在我历次去欧美国家的学术访问中，对我影响最深的也是宗教哲学和道德哲学。遗憾的是我未能抽出充裕的时间在这些方面下工夫。1988年夏我在美国进行学术访问期间，正在温德巴特大学攻读哲学博士的朱新民先生竭力向我介绍和推荐麦金太尔和他的《寻求德性》（一译《德性之后》，原文为“After Virtue”）。当时麦金太尔正要离开该校前往圣母大学任教。回国后我不断地向研究生们介绍这一情况。上世纪90年代初以来，我国学术界对麦金太尔的关注不断升温，他的主要著作逐渐被译成中文，尤其是《寻求德性》，引起学者们的浓厚兴趣，这部著作与罗尔斯的《正义论》成为近十几年来我国学术界的热门话题之一。对人生、对生活世界的关注，将形而上与形而下结合起来是20世纪西方哲学发展的重要动向，20世纪下半叶特别是六七十年代以来，这一倾向更加明显。西方学界越来越重视社会政治哲学、伦理学等“实践哲学”的成果。90年代以来，我国哲学界也越来越关注西方“实践哲

学”。学术界的这种动向显然与国际国内社会急剧变化、社会问题越来越突出有关。当然，伦理学、道德哲学的有影响的成果是极其丰富多样的，决不限于上述麦金太尔和罗尔斯的著作。要全面了解当代世界伦理道德思潮，必须扩展我们的视野。呈现在读者面前的这本由德里克·帕菲特所写的《理与人》，对中国学者来说恐怕还是比较陌生的。帕菲特其人其书在中国学界尚不像麦金太尔及其著作那样广为人知，但是在西方当代伦理学领域，帕菲特的地位和影响并不亚于麦金太尔，《理与人》一书在学术上的价值和意义是很高的。该书对传统伦理道德观特别是西方人在合理性、道德、个人同一性等方面的根深蒂固的信念提出了挑战，作者提出的一系列观点富有创见，在当代西方学术界自成一家，具有代表性。这本书的出版在西方伦理学界被视为学术上的一个重大的发展，给予很高的评价。当代著名哲学家斯特劳森在《纽约书评》中评价该书说："此类著作中鲜有在范围、丰富程度、想象力和推理的说服力方面可与之相媲美者。"另一个著名哲学家舍夫勒在《时代周刊》中认为："《理与人》可能是自百年前西季威克发表《伦理学方法》以来功利主义传统中有实质性的道德哲学的最伟大的著作。"还有人用"才华横溢"、"思想丰富"、"天才之作"等来评述《理与人》。美国《哲学论坛》1999年12月发表的一份关于20世纪西方哲学经典的调查报告将该书列入25本经典之中。该书在元伦理学和应用伦理学领域都有重要地位，它对当代西方特别是英美大学中的伦理学课程产生的影响更加显著，该书和麦金太尔的书同是当代伦理学课程的必读书。在为期一学年的当代伦理学课程教学中，这本书通常是第一学期的研读用书，第二学期则是中国学界较为熟悉的麦金太尔的著作。90年代初以来我国学术界越来越关注现当代西方伦理学，翻译介绍了大批西方伦理学著作，麦金太尔的著作和理论更是备受关注，而帕菲特的著作和理论却无人知晓，这不能不说是我国学术研究中的缺陷。我们翻译帕菲特的这本著作就是希望能填补我国学术研究中的空缺和不足，推动我国的学术研究和高校的学科建设。

二

正如书名所示，该书讨论的主要论题有两个：理（Reason）和人（Person），即行动的理由（道理）和行动者（人）的本性。作者认为，我们有理由去进行的行动，总是与行动者对人的本性的看法相关，因此，这两个主题是密切关联的。在现在的境况中，如果我的本性与我原先所认为的不同，我可能不得不修正有关我有理由去做的事的观念。我们关于我们（人）的本性的观念一直处于错误的状态，所以我们必须重新思考我们认为有理由去进行的观念。关于行动的理由，即"什么是我们最有理由做的事"这一问题，有两种解释的理论。一种是道德理由，即要道德地行动，这是道德理论；一种是关注行动的合理性，即要合理地行动，这是合理性理论。道德与合理性是密切相关的，作者是将两者连在一起讨论的。全书的前半部即第一、第二编主要讨论这方面问题，它以传统和流行的自利论与后果论为分析对象，以自利论（Self-interest Theory，以字母S代表）为合理性理论、以后果论（Consequentialism，以字母C代表）为道德理论，详细论证了在什么情况下它们并不导致直接地"自我挫败（Self-defeating）"，在什么情况下会直接地"自我挫败"。所谓"自我挫败"是指一种理论"即使以其自身的标准来衡量也是失败的"，也就是说，应用这种理论的结果达不到其目的。

自利论给每个人设立的目标是：对他自己来说结果要最好，而且会使他的生活过得对他自己而言尽可能地好。自利论的中心主张是，人人都有一个至高无上的合理的终极目标：使自己的生活过得对自身而言尽可能地好（《理与人》第5页）。当付诸行动的时候，自利论主张：我们每一个人最有理由做的事是对自身最好的任何事；任何一个人做他认为对自身来说将更糟的事是非理性的。在我们不能预测我们行动的后果的时候，自利论主张对任何一个人来说，做来合理的事

是将带给他最大预期利益的事。关于愿望和意向的合理性，自利论主张至高无上的合理的愿望是自己的生活对自身而言过得尽可能地好，至高无上的合理意向是某个永不克己的人的意向（第11页）。后果论的中心主张是，存在一个终极的道德目标：结果尽可能地好。后果论适用于一切事情。当应用于行动的时候，后果论主张下述两者：我们每一个人所应当做的是任何产生最好结果的事；如果某个人所做的是他相信将使结果更糟的事，他就是在错误地行动（第35—36页）。作者指出，自利论主宰我们的理智传统由来已久，两千年来自利论一直得到大多数西方人的信奉，原因有二：一是这些人中大多数相信来世或转世，因而认为道德论和自我利益总是相契合的；二是因为自利论的某些部分是有可信度的，如更多关心一个人自己并不是非理性的，关心自我利益时间上是中立的，等等。因此，如果迫不及待地主张我们应当拒斥自利论可能看起来有些荒唐。在这本书中，帕菲特的确不是简单地否定自利论，他从直接和间接、个体和集体的不同角度对自利论进行分析论证，不仅从理论上而且从实践层面上揭示和分析自利论所面临的难题，他以具体的事例和严密的逻辑论证，来说明自利论究竟是不是一种“自我挫败”的理论。在论证中他还将自利论与各种相关相对的理论进行比较，或者说从其他理论对自利论进行挑战，其中最主要的是道德论（Moral Theory，以字母M代表）和当前目标论（Present-aim Theory，以字母P代表），尤其是后者，它采取工具论（Instrumental Theory，以字母IP代表工具论的当前目标论）、审慎论（Deliberative Theory，以字母DP代表审慎论的当前目标论）和批判性（Critical，以字母CP代表批判性的当前目标论）三种形式（版本）。IP主张：我们每个人最有理由做的，是最好地实现其当前愿望的任何事。DP主张：我们每个人最有理由做的，是行动时已经经历“理想的审慎”过程，能够最好地达到他本会要的东西的事。CP主张：某些愿望是内在地非理性的，并不为行动提供好的理由，某些愿望是合理地要求的（第171页）。这三种形式中作者最支持的是批判性的

当前目标论。他指出，人们往往把当前目标论特别是审慎论与自利论看作是相契合的，认为如果每个人真正知道事实而且思虑清楚的话，那么他最想要的会是去做对他而言最好的任何事，或者会是最好地促进其长远的自我利益的任何事。这一假定被称作"心理自我主义"（第185页）。但实际上，一个人最想要的东西（愿望实现）与这个人的自我利益并不总是一致的，如某人最想要的是尽职尽责，但却往往有悖于自己的利益，即尽职尽责往往要牺牲自己的利益。人们实际想要的（愿望）与其利益之间严重背离的情况实在太多了。自利论与当前目标论经常相冲突。行动的时候最好地实现我们的各种愿望的东西与最有效地促进我们自己的长远利益的东西时常无法相契合。两千年来自利论一直占据着统治地位，基督徒们一直诉诸自利论，认定道德与自我利益是相契合的。其实不然，自利论既与当前目标论又与道德论相冲突。自利论是否合理的问题与时间有关，因为任何理由的力量都是在时间中展开来的，因此作者又通过各种事例仔细分析和论述了合理性与时间问题。结论是对两千年来大多数西方人相信的自利论持反对态度，即向传统观点提出挑战。作者指出，自利论并不仅仅是谨慎，而且是自我主义。它坚持的是，一个合理的行动者必须赋予他自己的自我利益以最至高无上的权重，他人付出无论什么代价都不足惜（第281页）。他还认为，"有关合理性的所有可能的理论都是批判性的当前目标论的一些版本"，无论什么是我们所相信的，"我们都应当接受批判性的当前目标论"（第283—284页）。

三

"合理性"、"道德"问题总与人相关，该书后半部分展开了关于人的问题的讨论，即第三编"个人同一性"（Personal Identity）和第四编"未来的世世代代"（Future Generations）。个人同一性也是西方哲学史上的老问题，洛克、休谟等都提出了自己的理论，帕菲特讨论了个

人同一性与合理性、与道德的问题，他依据现代科学对传统的问题和论点设计了许多富有想象力的事例和诘难。他认为，改变关于个人同一性的本性的观点，会影响我们关于合理性和道德的信念。帕菲特的论点是，个人同一性有两个类型。与此相关，存在着两个相互敌对的理论，即还原论者和非还原论者的观点。还原论者认为，一个人的存在就寓于一个大脑和躯体的存在之中，以及一系列互联的物理和精神事件的发生之中（第305页）。非还原论者认为，我们是单独存在的实体。照此观点，时间中个人的同一性并不就寓于物理的连续性和/或心理的连续性。它涉及一个进一步的事实。一个人是一个单独存在的实体，有别于他的大脑、躯体以及经验。这种观点的一个最为人所知的版本是认为，一个人是一个纯粹精神的实体，即一个笛卡儿式的自我或者精神实体。但是，我们也可能相信一个人是一个尚未在当代物理学理论中获得承认的那类单独存在的物理实体。所以，还有另一种非还原论者的观点。这种观点否定我们是有别于我们的大脑、躯体和经验的一些单独存在的实体，认为个人的同一性是一个进一步的事实，个人的同一性并不就寓于物理的连续性和/或心理的连续性之中。作者称之为进一步的事实观点（第305页）。帕菲特认为，我们有充足的理由拒斥非还原论者的观点，它们不论以什么样的形式出现，都是错误的。

为了支持他的论证，帕菲特设想了各式各样的匠心独具的事例和诘难。其中一个带有科幻色彩的例子是，在一个电子传输器上以光速把你送达火星，几分钟后你的精确复制品在火星上出现了。与此同时，你在地球上的躯体和大脑遭到了毁坏。这个复制品具有你的记忆、意图、习惯和情趣，他将贯彻你的计划、照料你的家庭等等。简言之，他在所有相关方面都与你是连续的。但是在什么意义上你还是那同一个人？帕菲特论证说，这种连续性所提供的就如同个人同一性一样，按照个人同一性的宽泛的标准“我的复制品……会是我”。但非还原论者会反对说，这种说明没有恰如其分地对待有关人的某些关键事

实。反对还原论方面的有价值的那些事实包括：人的"严格的和哲学的同一性"，它与一种"宽泛的和通俗的"同一性概念是不同的；"意识的统一性"，即认为我的各种各样的感知、思想、感情等等都只有一个主体；第一人称所指称的现象，有时要求代词"我"具有一个独特的所指，即自我。对于这些"事实"，帕菲特明确认为，在不设置任何严格的同一性观念的情况下，心理的和/或身体的连续性和联系性足以说明人的同一性，尽管有人对此持异议。他一方面承认意识的统一性所提出的问题对他而言构成有待进一步探讨的一定的麻烦，另一方面他效法利希滕贝格和威廉斯的论证，认为，尽管"在描述数不清的思想、愿望和其他经验的内容的时候某些个人是必定被提及的。但是……这样的一些描述并非主张这些经验是被某人所拥有的"（第325页）。总之，帕菲特指出："当从道德上考虑的时候，更少地聚焦于个人——经验的主体，代之以更多地聚焦于诸经验本身，变得似乎更加令人信服。就像我们忽略人们是否来自相同的还是不同的国家是正确的一样，我们忽略诸经验是否来自相同的还是不同的生命也是正确的，此点变得似乎更加令人信服。"（第488页）

个人同一性问题在西方哲学史上为哲学家们所关注，其中最典型的是洛克和休谟，本书也经常提到这两个人。洛克把意识特别是记忆看作是个人同一性的标准，忽视了肉体持续性。休谟认为，个人同一性概念是一个虚构，因为我们确实知道的惟一存在物是知觉，我们所感知的只是印象和观念，所谓心灵、自我不过是流变中的知觉的集合或一束知觉，我们把这些特殊的知觉归为一个自我是没有正当理由的。帕菲特对这些观点持批判态度，他试图采取比较灵活的立场，防止同一性问题上的僵硬观点。在他看来，对我们最重要的是我们生活（生命）中的连续性和联系性，和我们作为一个人的继续存活。他重视同一性问题是因为他认为，对个人同一性的看法会影响我们的情绪和我们关于合理性和道德的信念，也会影响到我们有关平等原则和其他一些分配原则的信念（第495页）。

四

该书的第四编“未来的世世代代”，讨论了道德责任应包括未来人的利益的问题。作者认定，我们对未来的世世代代的人负有重大的道德责任（义务），就此展开了错综复杂的充满功利主义色彩的论证。这些论证是从先前那些部分所得出的结论出发的，即从对自利论的拒斥和人的同一性的相对性理解出发的。他认为，如果我们造成污染，那么，即使在未来一些年不会影响到任何人，那也是坏的，而且在至关重要的意义上这就像我污染了我马上要在里面洗澡的水一样坏。这关涉未来人的利益问题。在这方面如果仅仅从时间上的区别或者人的区别的角度来看问题，都不能为道德评估提供依据。假设我向森林深处射出一箭，伤到了那里的某个人，我犯有疏忽罪。并不能够因为这个人离我太远，我不能辨认我所伤的是谁，就认为自己是无罪的。这个人离得太远不是无罪的理由，也不是逃避责任的理由。对在空间上离我们遥远的人，我们应如此主张，有关对于那些时间上遥远的人们的影响，我们应当持同样的主张（第510页）。作者分析论证了会影响未来世世代代的社会政策，如能源问题上的消耗政策和冒险政策；分析了未来的人口问题、生活质量问题等等。

总之，《理与人》以非常有力的论证动摇了西方人在“合理性”、“道德”和“个人同一性”等问题上的某些最根深蒂固的信念。作者认为，我们大多数人应当改变有关个人的本质和历时的个人同一性的观点；自利论和后果论是经常间接地自败的，违背我们的最大利益的行动时常是合理的，我们应当拒斥关于合理性的自利论；我们绝大多数人具有的道德观念（常识道德）是经常直接地集体性地自败的；在考虑我们的行动如何影响到他人方面，我们绝大多数人都犯了严重错误，因此，我们应当改变我们现在的许多行动方式（第637页）。当我们考虑到未来人的时候，我们难以避免一些令人不安的结论，我们

需要一个有关善行的新理论。在这些论断的基础上，作者又指出，非宗教的伦理哲学是一个新生事物，前途远大，但也不可预测。就西方文明的传统而言，伦理道德和宗教是相互连结、交织在一起难以分开的。所以帕菲特认为，人类的历史和伦理学的历史可能刚刚开始。因为在这之前，将非宗教伦理学作为毕生事业的人"为数极少"，在大多数文明中大多数人相信上帝存在或者多个神的存在，无神论者是少数派，非宗教伦理学得到许多人的系统研究，还只是20世纪60年代之后的事。与其他科学相比较，非宗教伦理学是最年轻也是最不发达的一门科学。不信上帝，得到多数人的公开容许，只是最近的一个事件，尚未完成。因为这个事件过于新近，所以非宗教伦理学尚处在一个非常早期的阶段。我们尚不能预言我们是否将会像在数学中那样在非宗教伦理学中达成共识（第644页）。

五

本书涉及许多问题，范围很广，包含大量经验的和设想的资料，内容丰富充实，逻辑推理和论证严密。作者将全书划分成154节，从内容到形式都表现出英国传统的经验主义和现代分析哲学的印记和背景。该书的翻译和阅读有相当高的难度，主要难点是：对书中的许多资料和论证不熟悉，我国学术界对该书讨论的许多问题缺乏研究，尤其是我们对道德哲学、伦理学的研究思路、思维方式与该书作者不一样。这些难点与我们在阅读西季威克的《伦理学方法》时所遇到的困难有类似之处。廖申白先生在《伦理学方法》译序中谈到阅读该书困难时所引用的一段话，同样适用于《理与人》："他（西季威克）不断地提炼、限定，提出反对意见，回答，再提出进一步的反对意见。所有这些反对意见、回答、反驳、再驳本身都是令人起敬的，并且也的确表现了作者的敏锐和坦率。但是读者却容易失去耐心，弄不清论据之所在，甚至会推案而起，觉得他满怀敬意地阅读了那么久，对书中

的内容却毫无记忆或所记甚少……”我们在译、校《理与人》的过程中对这类难处深有体会。因此，我们在这里所作的介绍决不是全面的，更谈不上专门研究。西季威克是英国维多利亚时代的著名哲学家，他的《伦理学方法》是继古典功利主义大师边沁、穆勒之后19世纪西方最重要的伦理学名著之一。帕菲特在《理与人》中多处提到这本名著。尽管帕菲特并不完全赞同西季威克的观点，但西季威克的影响却清晰可见，无论是在思路、观点方面，还是在具体论述方面。当然，这两本著作在时间上差了整整一个多世纪，这是一个历史快速发展的时期，相比之下，《理与人》与现实生活的关系更加密切，更为鲜明地反映了社会经济、政治、科技发展中的矛盾冲突和问题，诸如人口、资源、环境、核武器等等所蕴含或引发的伦理道德问题。

我国的传统文化蕴含着丰富的伦理道德思想，但是，其中缺乏与现代化和市场经济相适应的内容，这是因为我们的现代化进程较之西方要晚得多，市场经济刚刚起步。但是，晚近二十多年来，随着全球化进程的加快，特别是我国经济和社会的高速发展，我们面临着许多与西方发达国家相类似的问题，伦理道德是其中较为突出的一个。当然，《理与人》并没有穷尽这些问题，更没有提供完善的解决方案，许多观点也不是我们所能赞同和接受的，然而其中确实也不乏出色的、有价值的、有启发的思想、见解、论述，有不少的闪光点。比如，作者指出，我们大多数人现在生活在一些大共同体中，我们行动的糟糕后果现在可能殃及数以千计，甚至数以百万计的人们；但是，如果一个人的行动伤害到他人而其影响微不足道或未被察觉的话，我们就会认为这个行动并非失当。他严肃地指出，这是非常普遍而十分有害的错谬的信念，施虐者就利用这一点明知故犯地实施暴虐，这样的事例很多。帕菲特呼吁我们纠正这种错谬信念，将这类行动看作在道德上是错误的。因为尽管一个行动的后果微不足道，但是，合起来会极大地伤害我们自己或他人。众多微不足道未被察觉的行动后果积聚起来就会造成严重后果，这方面的例子有：污染、拥挤、损耗、通

货膨胀、失业、经济衰退、过度捕捞、过度耕种、水土流失、食物匮乏和人口过剩（第628页）。在日常生活中我们对一些有损公益的行动往往因为其后果不显著而不予重视，不加谴责，不予制止。然而，正是这些"微不足道"的行动后果日积月累酿成灾难。帕菲特的经验洞察力是很强的，这一见解是很深刻和有价值的。

在英美国家，伦理道德的教学和研究崇尚推理形式，《理与人》就是一部典型的道德推理著作。它没有说教，不颁布训条，不发号施令，不以无可置疑的真理自居，不向读者灌输先定的信念，不提供现成的确定的答案。总之，它不是独断论的，而是批判论的，它对传统的现成的理论和论断置疑和挑战，贯穿全书的是讨论、诘难、辨析、比较，不固执于一方一词，从各种不同的乃至对立的立场、观点、视角出发，进行推理、辩驳，引出结论。读这本书会使人想起苏格拉底的论辩术，也会想到中世纪经院派唯名论唯实论哲学家的逻辑论证。但是，本书的道德推理并不是脱离实际的空洞的文字游戏，作者立足于各种各样的经验事例，又常常超越经验，依据现代科技，发挥奇妙的想象力，设计出充满科幻色彩的趣味盎然的事例。这样的道德推理无疑可以砥砺我们的思维能力，但同时又使我们感到困惑不适应，在阅读这些逻辑论证时，我们常常感到仿佛置身于一座迷宫之中，因为这种道德推理与我们传统的习惯的思维方式和研究思路大相径庭。我们无意在这里对道德推理的优劣得失作出全面的评论，但毋庸置疑的是，我们应当重视道德推理，以此来充实我们的伦理学研究。长期以来，我们对于在英美国家如此盛行的这种道德理论和方法漠然处置，这是不应该的。当今时代，我国的伦理学研究不仅具有自己独特的问题域，而且面临着全球性的共同的和相似的问题系，我们应当拓展我们的视野和思路，防止思维方式的单一和僵化。

翻译外国学术著作是严肃而又重要的工作，是学术研究不可或缺的基本建设之一。本书由王新生译、黄颂杰校。谨以此书献给复旦大

学百年校庆。

由于学识浅陋，水平有限，时间紧迫，本书在译、校方面必定存在许多错误不当之处，恳请读者批评指正。希望本书能为我国的学术研究提供有用的参考。

黄颂杰

引论

我常常像我的猫一样，想怎么做就怎么做。而且此时我并不是在使用一种只有人才具有的能力。我们知道，行动的理由有多种，有些理由比其他一些更好更充分些。本书的主要论题之一便是关于我们有理由去做什么的问题集。我将讨论好几种理论，其中有些是道德理论，其余的是关于合理性的理论。

我们每个人都是独特的人。我有我的生活方式，你有你的生活方式。这些事实意味着什么？使我一生都是同一个人、一个与你不同的人的东西是什么？这些事实的重要意义是什么？每个人生的统一、不同生活和不同人之间的区别，它们的重要意义又是什么？这些问题是本书的另一个主题。

本书的两个主题理与人是密切相连的。我认为，大多数人对于人的本性和人生的同一性有着一些错误的信念。因此，一旦我们明白真相，我们就应当改变某些关于我们有理由去做什么的信念。我们应当修正我们的道德理论和我们关于合理性的信念。在本书的前两编中我还对其他类似的结论作出了论证。

我并不想预先叙述这些论证和结论，本书的目录提供了一个概要。本书篇幅很长，有些地方也比较复杂。因此我将我的论证分成154节，每一节都有一个描述性的标题。我希望这样做会使这些论证更容易得到理解，并且能够比主题索引更为清晰地表明本书所包含的内容。如果我不把这些论证一节一节地区分开来的话，那么，即使编写一份主题索引，也可能会因为这份索引太繁杂而无多大用处。

同类书的许多引论都试图解释在这些书中所使用的一些核心概念。但由于这样的解释要想有助于读者至少要花上一本书的篇幅才行，因此我就不想耗费时间来这样做。我书中的核心概念并不多。我

们具有行动的理由。我们应当按某些方式去行动，而某些行动方式在道德上是错误的。某些结果就其与道德相关而言是好的或坏的：例如，假如一个人变成瘫痪，它是件坏事；而如果可能的话，我们就应当去阻止它发生。大多数人想必能理解我最后所讲的三句话，如此就足以理解我的论证了。我将使用的概念还涉及某人的自利包括什么，或者什么东西是对此人最好的。我在附录A中对此作了简要的讨论。本书最后一个核心概念是人。大多数人认为，我们已理解了人是什么。本书第三编则认为，对此我们并不理解。

许多书的引论还试图说明，当讨论道德时我们怎样有望从中取得进展。既然最好的说明就在于它确实在取得进展，那么，这就是我将努力作出的惟一的说明了。

斯特劳森讲述有两类哲学：描述的和修正的。描述的哲学为我们凭直觉所假定的东西提出理由，并且为我们的关于我们自身和我们居住于其中的世界的那些信念之中的不变的重要内核进行解释和辩护。我十分尊重描述的哲学。但是，气质决定了我是一个修正论者。在本书乏味的第一章中，我无法避免重复那些已被表明为真的东西，除此之外，我试图向我们所假设的那些东西进行挑战。哲学家不仅应当解释我们的信念，当信念虚假时，我们也应当改变这种信念。

1985年重印注释：这次重印时我作了某些修改。我不再支持个人同一性的宽泛心理准则（第208页和别处[1]），因为这一看法与我另一观点（即我们不应试图在不同标准之间作出决定）相冲突。我在声言Q（第360页）和后面几句中用“更坏”取代“坏”。其他较小而有实质性意义的修改包括第一编注释[2]、第三编注释[3]和第66页15—17行、67页19—23行、211页19—25行、312页31—32行、374页1—2行和12—13行。其中有些修改我要感谢《伦理学》（1986年7月）的撰稿者们。

1987年重印注释：这次重印我放宽了还原论（归约论）的定义（第210页），并且把明显的循环性从物理和心理准则中去除。其他较小而有实质性的修改是第72页2—4行、73页5—15行、81页17—24行、82页5—11行、311页40行和第一编注释[4]。（我还作了某些文体上的修改。）

---

[1]本文出现的页码均为原文页码。——译者

[2]（此注释为1987年所加）我所举的例子有缺陷。对于我而言造成我自己失去这个意向并不会是错误的，如果我做到这一点是通过以我的敌人毁掉我的生涯为结果来拒绝帮助我的敌人的话。这个例子可以加以修补。（我们可能假设，他只会损毁我的生涯，因而能威胁做出进一步损害。）

[3]（该注释为1985年所加）正如B·加雷特（B.Garrett）已经指出的那样，此节有错。就亲缘而论，重要的不是"……的亲属"的传递关系，而是能够具有程度"……的近亲"之非传递关系。如果我们假定重要的是个人同一性，那么这因而就不是一个好的类比。要回答该反论证，我们可能不得不挑战这个假定。

[4]（此注为1987年所加）在其刊载于1986年7月《伦理学》之中的那篇文章里，A·库夫里克论证说，常识道德并不包括我所称作M的东西的整体，因而并不是自败的。这可能如此。但是我的主张本来会是，假如常识道德的确包含了M的整体的话，它就会是自败的。更重要的是，除非这个道德包含我的修正R，否则，就将有许多协同难题，对此它令人不快地未能解决。（参见同一期中我的"回应"。）


第一编 自败的理论
上海译文出版社

第一章

间接自败的理论

我们最有理由做的是什么？有多个理论可以回答这个问题。其中一些是道德理论，另一些是有关合理性的理论。一旦应用于我们的讨论，不同的理论会给予我们不同的回答。那么我们必须努力判定哪一个是最好的理论。

有关这些理论的论证种类繁多。其中之一是说有的理论是自败的。这种论证不需要什么假设。它主张，有的理论即使以其自身的标准来衡量也是失败的，从而宣告自己完全无用。

本书第一编讨论的是这种论证所取得的成就。正如我将要解释的，所有那些最为有名的理论在某些方面都是自败的。这表明什么？在一些情况下，什么也没有表明。在另一些情况下，所表明的是一种理论必须加以进一步发展或者展开。还有一些情况下，所表明的是一种理论要么必须加以拒斥要么必须加以修正。这就是有关我们绝大多数人所接受的那些道德理论所表明的。

我就从最有名的那种理论开始。

1.自利的理论

我们可以通过说出理论告诉我们所力求达到的是什么来表述所有的理论。根据所有的道德理论，我们的行动应当力求符合道德。根据所有关于合理性的理论，我们的行动应当力求符合理性。我们把这些称为形式上的目标。而不同的道德理论和不同的关于合理性的理论给我们设立了不同的实质性的目标。

我用"目标"一词意指"实质性的目标"。目标一词的这种用法是广义的用法。它能够表述那些并非关于道德目的，而是关于权利或者义务的理论。按照某个理论，假设有五种行动是完全被禁止的。这个理论就给我们设定了永远不要以这五种方式去行动的目标。

我将先讨论自利论（或用缩写S表示）。这是一种有关合理性的理论。S给我们每一个人设立了这样一个目标：对他自己来说结果要最好，而且会使他的生活过得对他自己而言尽可能地好。

要想应用S，我们必须问及这个目标最好会达到什么。对这个问题的那些解答我称作关于自利的理论。正如本书“附录”A所解释的，令人信服的相关理论有三个。

按照快乐论，对某个人来说最好的东西就是会带给他最大幸福的东西。至于幸福牵涉到什么和幸福如何度量，这个理论的不同版本则持有不同的主张。

按照愿望实现论，对某个人来说最好的东西就是一生中最满足他愿望的东西。在此，这个理论同样存在不同的版本。如此这般，成功论诉诸的就只是某个人有关自己生活的那些愿望而已。

根据目标列表理论，即便我们不想拥有好的事物或者不想规避坏的事物，某些事物对于我们来说也仍然要么是好的要么是坏的。在此，这个理论同样具有不同的版本。那些好的东西或许包括能力的发展、知识的进步、审美的增强。那些坏的东西或许包括施虐的快感、受骗和失去自由或尊严。

这三个理论部分重叠。按照所有这些理论，幸福和快乐至少是那种使我们的生活变得更好的一部分原因，悲惨和痛苦至少是那种使我们的生活更糟的一部分原因。这些主张会出自任何一个令人信服的目标列表理论，而且为愿望实现论的所有版本所蕴含。按照所有的理论，快乐论至少是真理的部分。为节省笔墨起见，有时这就是我予以讨论的惟一部分。

所有这些理论还主张，在决定对某人来说最好的会是什么的时候，我们应当给予这个人未来的所有部分以同等权重。较后一些的事件可预测性较少；如果一个可预测的事件发生的可能性较少的话，那么重要性就应当算得较少一些。但是不应当只是因为如果它发生它会较后发生而把其重要性算得较少一些。

在那些关于自利的不同理论之间作出决断至少会占用整本书的篇幅。本书讨论这些理论之间的一些差异，但是并不试图在它们之间作出决断。本书的相当篇幅讨论自利论。正如我说过的，这不是那些关于自利的理论中的一个。它是关于合理性的一个理论。我们可以在对自利的不同理论不作出决断的情况下讨论S。我们可以提出按照所有这些理论来说都会为真的一些主张。

把一些目标称作终极性目标将有所助益。其他一些目标则是工具性目标，纯粹是一些达到某个终极性目标的手段。这样说来，对于除了守财奴之外的所有人而言，富有并不是一个终极目标。我现在可以重新表述自利论（S）的主要主张。

这就是：

（S1）人人都有一个至高无上的合理的终极目标：使自己的生活过得对自己而言尽可能地好。

正如我们将会看到的，S还提出了数个别的主张。

对S有好几个诘难。其中的一些我在第二编和第三编中予以讨论。下面我讨论的是这样一个诘难，即，S像某些别的理论一样是自败的。

2.自利论何以可能是间接自败的

假使我们称某个理论为T，称其给我们所设定的目标为我们的T设目标。在以下情况下，我们称T

间接地、个体地自败的，如果当某人力图达到他的那些T设目标时，这些目标的达成总体上将更糟这一点为真。

按照这个定义，我们并不只是简单地问一个理论是否自败。我们问它在某个时期应用于某些人的时候是否自败。

我的S设目标是，让我的生活对我而言过得尽可能地好。如果我力图做任何对我而言最好的事情，对我而言反而都将变得更糟，这一点能够为真，有两类情形：

（a）我的那些尝试可能时常失败。我可能经常做出一些比其他我会做的什么对我而言更糟的事。

（b）在我可能做出的那些行动中，即便我从不做那种对于我而言将会更糟的事，但是如果我是纯粹自利的，那么对于我而言也可能是更糟的。如果我具有某种别的意向的话，对我而言或许更好。

在那些（a）类情形中，糟糕的后果来自我的所作所为。假设每当我相信自己不会被抓住的时候我行窃。我就可能时常被捉住，遭到惩罚。即便从自利的角度来说，诚实因而可能是对我而言最好的策略。这些情形并不值得讨论。如果这就是S自败的方式，那么这根本就不算是对S的什么诘难。S在此是自败的，只是因为我在试图遵循S方面的无能。这是一个缺陷，但并非S的缺陷，而是我的缺陷。我们或许诘难某个理论过于困难以至于无法遵循。但是这一点对S却并不为真。

值得讨论的是（b）类情形。在这些情形中，即便我成功地做到从不做那些对我来说更糟的事，如果我纯粹自利的话，对我来说也将更糟。糟糕的后果不是来自我的所作所为，而是来自我的意向，或者说来自我纯粹自利这一事实。

这个事实涉及什么？我可能会是纯粹自利的，但并不是完全自私的。假设我爱我的家庭和朋友。按照所有有关自利的理论，我对这些人的爱影响到符合我的利益的东西。我的相当多的幸福来自于知道那些我所爱的人是幸福的，以及有助于带给我所爱的人以幸福。按照愿望实现论，假设如我所愿，我所爱的那些人一切顺利，这对我来说更好一些。对我来说最好的东西，在各个方面，与对我所爱的那些人来说最好的东西很大程度上是重合的。但是，在有些情况下，对我来说

将会更好的东西对我所爱的那些人来说将会是更糟。如果我在所有这些情况下都做对我而言将会更好的事情，我就是自利的。

可能有这样的想法，即，如果我是自利的，我将总是在力图做任何对我来说最好的事情。但是会有这样的情况，明知两种方式中没有哪一种对我来说会更好一些，我还是时常以两种方式中的一种方式行动。在这些情况下，我并非是在力图做对我来说最好的事情；我是在按照一个更为特定的愿望行动。甚至我在做知道对我来说将会是最好的事情时，这一点也可能属实。假设我知道如果我帮助你的话对我来说将是最好的。我之所以可能帮助你，是因为我爱你，而不是因为我想做一些对我来说将是最好的事情。在表述我要是自利会是什么样子的时候，这样主张就足够了：尽管我时常按照别的一些愿望在行动，但是我从不做我认为对我来说将更糟的事情。如果这一点为真，那么就不是称我自利，而是称我从不克己，恐怕更加明确一些。

我现在将重新表述那个对任何个体而言S可能间接自败的有趣方式。下述情形就是如此：如果某个人从不克己，这对他而言将比他倘若有某种别的意向的情况更糟。即便某人成功地做到从不做那种对他而言会更糟的事情，但是他从不克己对他而言可能会更糟。如果他有某种别的意向的话，或许对他更好一些。因此他有时或许会做对他来说更糟的事情。但是对他来说以这种方式行动所付出的代价或许少于具有这种别的意向所带来的利益。

这些主张按照所有关于自利的那些不同理论都能够为真。快乐主义者早就知道，当以幸福为目标的时候，幸福反而更加难以达到。如果我最强烈的愿望是我幸福，我可能比我有其他更加强烈的愿望的话本来会有的幸福更少。如此说来，如果我最强烈的愿望是别的什么人的幸福的话，我或许会更加幸福。

再举一个例子。凯特是一位作家。她最强烈的愿望是使她的著作尽可能地完美。由于她如此在意她的著作的质量，她发觉她的工作颇有回报。如果她写好书的愿望弱一些的话，她会发现她的工作相当乏

味。她深知这一点，而且采纳了关于自利的快乐论。她因而相信，如果她的最强烈的愿望是她的书尽可能地完美的话，对她来说会更好。但是，由于这个愿望十分强烈，她通常工作得如此卖力，时间如此之长，以至于累垮了身体，而且在一段时期内相当郁闷。

假设凯特真的相信，如果她工作不那么卖力的话，她的著作会略微糟糕一点，但是她会更幸福。她会发觉她的工作同样颇有回报，而且会避免这些严重的郁闷。凯特因而相信，当她如此努力地工作的时候，却是在做对她来说更糟的事情。但是她为何从不以这种方式行动呢？事实可能是，只有在当她想使她的书尽可能完美的愿望弱得多的时候，她才不会以这种方式行动。但这对她来说甚至可能更糟，因为那样她会发觉自己的工作索然无味。按照快乐论，如果凯特从不克己的话，对她来说会更糟。[1]

假设我们采纳的不是快乐论，而是有关自利的愿望实现论。那么我们可能否认，在这个例子中，凯特是在做对她而言更糟的事情。她最强烈的愿望是使她的书尽可能地写好。通过如此努力地工作，尽管她把自己搞得精疲力竭而且相当压抑，但是她使她的书略微更好一些。她从而使她的最强烈的愿望得到更好一些的满足。按照我们有关自利的理论，这对她可能更好。

如果我们不是快乐主义者，我们就需要一个与之不同的例子。假设我在午夜驾车横穿某个沙漠。我的车出了故障。你是一位陌生人，也是附近惟一的另一位驾驶员。我设法把你拦下，而且我向你提出，如果你解救我的话，我会给予重酬。我现在不能够酬谢你，但是我允诺当我回到家时酬谢你。接着假设我是坦率的，不会蒙骗别人。我不能够令人信服地说谎。无论是脸上的一阵泛红，还是我说话的语调，总会令我暴露无遗。最后假设我知道自己从不克己。如果你开车送我回家，如果我给你我所允诺的酬谢的话，对我会更糟。既然我知道我从不做那种对我来说更糟的事情，那么我知道我将不会遵守诺言。鉴于我没有令人信服地说谎的能力，你对这一点也心知肚明。你不相信

我的允诺，因而把我留在沙漠里继续受困。这之所以发生在我身上，是因为我从不克己。如果我值得信任，甚至在遵守诺言对我而言会更糟的情况下也愿意这样做，对我来说本来会更好。那样你就会解救我了。

（有一点可能受到诘难：即便我从不克己，但是我会决定遵守诺言，因为作出这样的决定对我来说将更好。如果我决定遵守诺言的话，你就会信任我，而且会解救我。这个诘难能够予以答复。我知道，在你驾车送我回家之后，如果我给你我所承诺的酬谢的话，对我会更糟。如果我知道我从不克己，那么我知道我将不会遵守诺言。而且，如果我知道这一点，我不会决定去遵守诺言。我不会决定去做我知道自己将不会做的事。如果我能够决定遵守诺言的话，这必定是因为我相信我将不会永远不克己。而且我们能够加上这样一个假设：除非它为真，否则我不会相信这一点。那么，如果我是而且将仍然保持从不克己的话，对我会更糟这一点就会是真的。如果我值得信任的话，对我会更好。）

我已经表述了倘若某人从来不克己的话对他来说会更糟的两种方式。还有许多别的方式也能说明这种情况。大多数人生活中的大部分时间很可能如此。当自利论应用于这些人的时候，它就是我所称的间接地个体自败。以其自身的标准来衡量S是失败的吗？S宣告自身无用吗？这取决于S是否告诉人们要永不克己。

3.自利论告诉我们要永不克己吗?

看上去似乎很显然，自利论（S）告诉每一个人要永不克己。但是正如到目前所表述的，S只是主张说，人人都有一个至高无上的、合理的终极目标：使自己的生活对自身而言过得尽可能地好。

当S应用于行动的时候，主张：

（S2） 我们每一个人最有[2]理由做的事是对自身最好的任何事，而且

（S3） 任何一个人做他认为对自身来说将更糟的事是非理性的。

S还必须就当我们不能预测我们行动的后果时应当做什么提出主张。我们可以忽略那些带有测不准性的情形，其中我们对不同后果的发生概率没有什么信念。但在那些冒险性的情形中，我们的确有这样的一些信念，S主张

（S4） 对任何人来说，做来合理的事是将带给他最大预期利益的事。

为要计算某个行动的预期利益，我们累加那些可能的利益，减去可能的代价，然后每一利益或代价乘以该行动产生利益或成本的几率。例如，如果某个行动带给我利益B的几率是9／10，带给我损失B两倍大小的某个利益的几率是1／10，那么预期的利益就是*B×9／10 - 2B×1／10*，或者说是B的7／10。

关于愿望和意向的合理性，S应当主张什么？既然S主张，每一个人都有一个至高无上的合理的终极目标，那么，S应当清楚明白地主张至高无上的合理的愿望就是那个达到这个目标的愿望。S应当主张

（S5） 至高无上的合理的愿望是自己的生活对自身而言过得尽可能地好。

同样，S应当主张

（S6） 至高无上的合理的意向是某个永不克己的人的意向。

如果某个人是从不克己的，尽管他有时候按照别的愿望行动，但是他从不违背那个至高无上的合理的愿望而行动。他从不做他认为对他而言将更糟的事。

为了简便起见，称愿望和意向这两者为动机。随着时间的变化，有一些我们能够使我们的动机改变的方式。我们能够养成一些习惯。

如果我们现在并不喜欢我们行动的那些方式，但我们可能最终会对它们喜欢起来。如果我们改变我们的工作，或者改变我们的居住地点，或者阅读某些著作，或者生养子女，所有这些都可能造成我们的动机中出现可以预测的变化。我们能够造成这样的一些变化的方式还有其他许多。

根据（S2），每一个人最有理由做的事是使他自己具有或者任由他自己保持以利己的角度看来最好的一组组可能的动机。这些是下述情况为真的那些动机：不存在别的可能的动机，使得如果这个人具有这些动机的话，会对他更好。我用"可能"意指"人类本质的一般事实和这个人的本质的特殊事实给定的情况下的因果上的可能"。某组动机对某个人而言是否会是因果上可能的，或者是否对这个人而言以利己的角度看来会是一组最好的动机之一，这些经常都难以知晓。但是我们可以忽略掉这些困难。在许多情况下某人的确知道，如果他的动机有所改变的话，对他会更好一些。而且在许多这种情况下，这样的一个人知道，以上述那些方式之一行动，他会造成这种变化。（S3）蕴含着，就这个人而言他不造成这种变化会是非理性的。

类似的主张适用于我们的情绪、信念、能力、我们的发色、我们的居住地以及其他任何我们可以改变的东西。我们每一个人最有理由做的是做出任何对自己而言最好的改变。如果某一个人相信他能做出这样的改变，对他来说不去这样做就会是非理性的。

我们现在可以回到我早先的那个问题了。我们是在讨论这样的一些人，对于这些人来说下述情况为真：如果他们从不克己的话，对他们而言这会比倘若他们有别的意向的话更糟。假设这些人知道这是真的。S告诉他们要永不克己吗？

S主张如下：如果这样的一个人是从不克己的，他会具有至高无上的合理的意向。但是对这个人来说，使自己具有或者保持这个意向会是非理性的。他使自己具有或者保持别的意向会是合乎理性的，因为这对他会更好。

这些主张可能看起来对我的问题给出了相互矛盾的回答。它们可能看起来既告诉这个人要永不克己，又告诉这个人不要永不克己。

这会误解S。当S主张说某个意向是至高无上地合理的时候，它并非告诉我们要抱有这个意向。不要忘记形式上的和实质性的目标之间的区分。正如所有关于合理性的理论那样，S给予每个人以这种形式的目标：合乎理性和合理地行动。不同的理论给我们设定了不同的实质性的目标，这一点使它们彼此区分开来。在其中心主张（S1）中，S给我们设定了一个实质性的目标：让自己的生活对自身而言过得尽可能地好。S给每一个人设定了另一个实质性的目标——合乎理性和合理地行动——吗？没有。根据S，我们的形式上的目标不是一个实质性的目标。

可能会有这样的想法，即，在提出这些主张时，我并未表述自利论的最好版本。而这个版本会被大多数相信这个理论的人所接受。他们大多数会采纳（S2）和（S3）。假设我知道，如果我使自己非理性的话，对我来说将是最好的。我将很快表述此点几乎肯定为真的一个情形。如果这点为真，（S2）就蕴含着我最有理由做的事情是去造成自己非理性。（S3）蕴含着我不去这样做会是非理性的。这些主张并未把合乎理性设定为我的一个实质性的目标。

这意味着对S而言合乎理性只是一种手段吗？这取决于关于自利的最好理论是什么。按照快乐论，S给每一个人设定了这样一个实质性的目标：对自身而言最大可能的幸福。合乎理性这一点并非这个目标的本质部分。它仅仅是一种手段。合理地行动、具有合理的愿望或意向也是如此。下面考虑目标列表理论。按照这个理论的某个版本，合乎理性这一点是那些对每一个人而言都好的东西中的一个，无论其后果如何。如果是这样，合乎理性这一点并非仅仅是一种手段，而是S给每一个人所设定的那个实质性目标的一部分。按照愿望实现论，在那些人们都想要合乎理性而无论其后果如何的情况下，这同样会为真。

它可能遭到这样的诘难：“假设我们采纳快乐论。那么S告诉我们，合乎理性这一点仅仅是一种手段。如果是这样，我们何必力求合乎理性？我们何必要知道我们最有理由做的是什么？如果我们接受S所主张的，而且认为合乎理性仅仅是一种手段，我们将不再对S主张要予以回答的那个问题感兴趣。这必定是对S的一个诘难。一个可以接受的有关合理性的理论不能够主张合乎理性仅仅是一种手段。”

我们可以如此回复：“如果一个理论主张合乎理性并不重要，它就会是一个无法接受的理论。但S并非如此主张。假设，作为逃生的一个纯粹手段，我攀附在某块岩石上。尽管我的行动仅仅是一种手段，但却是非常重要的。合乎理性这一点同样可能如此。”这可能没有完全回答这个诘难。正如我们将要看到的，对某些道德理论都有一个类似的诘难。为简便起见，我把这些诘难放到第19节中一并予以讨论。

我现在能够对我上文的一个说法加以解释了。根据S，那种至高无上的合理的意向是某个永不克己的人的意向。我上文中写到过，在提出这个主张时，S并未告诉我们要抱有这个意向。S给每一个人设定了一个实质性的目标：自己的生活对自身而言要过得尽可能地好。按照一些有关自利的理论，合乎理性这一点对一些人而言会是这个目标的一部分。但这只会是因为，就像幸福一样，合乎理性这一点是那使我们的生活过得更好的东西的组成部分。合乎理性本身不是一个实质性的目标。具有那个至高无上的合理的意向也不是。

就某些人来说，根据S，合乎理性这一点不会是使他们的生活过得更好的东西的组成部分。这些正是我们在讨论的那些人。就这些人来说的确是，如果他们从不克己的话，这会比具有某种别的意向对他们更糟。既然真的如此，从不克己这一点不会是S给这些人所设定的目标中的一部分。S并未告诉这些人要具有S所主张的那种至高无上的合理的意向：某个永不克己的人的那种意向。而且，如果他们能够改变他们的意向，S告诉这些人，如果他们能够，不要永不克己。既然如果这些人具有某种别的意向对他们会更好，S告诉他们要使他们自

己具有或者保持这种别的意向。如果他们知道能够以这两种方式中的任何一种方式行动的话，S主张说，他们不去这样做会是非理性的。他们使自己成为或者任由自己仍然保持永不克己会是非理性的。

4.以其自身的标准衡量何以自利论并未失败

这些主张回答了我所问的另一个问题。当自利论（S）应用于这些人的时候，它就是我所说的间接地自败。以S自身的标准衡量使它失败了吗？S宣告自己不适用了吗？

答案是否定的。S是间接自败的，因为对这些人而言如果他们从不克己则会更糟。但是S并非告诉这些人要永不克己，而是告诉他们，如果他们能够的话，不要永不克己。如果这些人永不克己，对他们将更糟。以S的标准衡量，这是一种糟糕的后果。但是这个糟糕的后果，既不是他们进行S告诉他们应当进行之事的结果，也不是抱有S告诉他们应当抱有的意向的结果。既然如此，S以其自身标准衡量并不失败。

有可能遭遇到这样的诘难："这个糟糕的后果可能是这些人对S的信念的结果。如果他们相信S，那么他们就会相信，去做他们认为对他们将更糟的事会是非理性的。可能为真的是，如果他们认为以这种方式行动是非理性的，他们永远不会去这么做。如果他们永不以这种方式行动，他们就是永不克己的。假设，在你所表述的那些方式的一个方式中，抱有这个意向对他们来说更糟。这是一个以S本身的标准衡量的糟糕后果。如果对S的信念有这个后果，那么以其自身的标准衡量S就的确是失败的。"

在回答这个诘难方面，我们需要知道这些人是否能够改变他们的意向。首先假设他们不能。这何以会成为真？如果他们不能够改变他们的意向，而且他们抱有这个意向是因为他们相信S，那么解释必定是：他们无法使自己乐意做他们相信是非理性的事情。只有当他们相信某个别的有关合理性的理论的时候，他们才会改变他们的意向。那

么S会告诉他们，如果他们能够的话，要使自己相信这个别的理论。这种可能性我在第6节到第8节中予以讨论。正如我们将要论证的，即便这一点是真的，以其自身的标准衡量S也不会是失败的。

其次假设这些人不经改变他们有关合理性的信念就能够改变他们的意向。如果这些人永不克己，这对他们来说将比他们本来抱有某种别的意向更糟。S告诉这些人要使自己抱有这种别的意向。上面所给出的诘难明显失败。可能为真的是，这些人永不克己是因为他们相信S。但是S主张，这些人任由自己一直保持永不克己是非理性的。如果他们确实一直保持永不克己，这就不能仅仅说成是一个"他们相信S的结果"。它是他们未能做本来能够做的事情和未能做S所告诉他们要做的事情的结果。这个结果对他们来说更糟，这是以S的标准来衡量的一个糟糕后果。但是一个其根源在于不遵守S的糟糕后果不能够造成对S的一种诘难。如果我的医生告诉我要搬迁到一个气候更有利于健康的地区，如果因为我拒绝搬迁而终究死亡，他就不会遭受什么批评。

还有第三种可能性。这些人可能既不能够改变他们的意向也不能够改变他们有关合理性的信念。他们对S的信念对他们而言是糟糕的，这是一个以S的标准来衡量的糟糕后果。这是对S的一个诘难吗？在我讨论过一些别的理论之后，这个问题将更容易回答一些。我的回答在第18节。

5.使自己非理性地行动能是合乎理性的吗？

我现在转到一个新问题。一种理论即便以其自身标准衡量并不失败，它仍然可能是不可接受的。就许多人的情况而言，如果他们从不克己，对他们的确会更糟。这能带给我们一些独立根据来拒斥S吗？

根据S，这些人各使自己抱有或者保持以自利的标准衡量最好的那些可能的动机组之一会是合乎理性的。至于是这些动机中的哪一个，在某种程度上这是一个事实问题。而且答案细节会因不同环境下

不同的人而不同。但是关于这些人中的每一个人我们知道下述情况。既然如果一个人从不克己对他会更糟，那么如果他有时克己的话对他会更好。如果他有时乐意做他相信将对他更糟的事情的话，那么对他则会更好。S主张，以这种方式行动是非理性的。如果这样的一个人相信S，那么它告诉他要促成自己乐意以S宣称为非理性的一种方式行动。这是一个破坏性的暗示吗？它会带给我们对S加以拒斥的任何理由吗？

考虑一下：

谢林对武装抢劫的回应。一个人闯入我的屋子。他听到我在报警。但是由于靠得最近的城镇也相当远，警察在15分钟之内不能赶到。这个人命令我打开我藏有黄金的保险箱。他威胁说，除非他在随后5分钟之内得到那些黄金，否则他就会一个接着一个地枪杀我的子女。

对我来说要做的合理的事是什么？我需要快速作出反应。我意识到把黄金给这个人并非是理性的。这个人知道，如果他只是拿走黄金的话，我或者我的子女都会把他驾着逃跑的那辆车的牌子和牌号告诉赶来的警察。所以存在这样一个巨大风险，就是如果他得到黄金的话，他驾车逃逸前会开枪杀了我和我的孩子们。

既然把黄金给这个人并非是理性的，那么我应当无视这个人的威胁吗？这也不是理性的。因为一种巨大的风险是，他会杀死我的一个孩子，促使我相信他的威胁：除非他得到黄金，否则会杀死我的其他孩子。

我应当怎么做？非常可能的是，无论我是否把黄金给这个人，他都将杀死我们大家。我处于一种绝望的境地。幸好，我记起读过谢林的《冲突战略》。[3]我手边还有一种特殊药物。这种药物能够导致人短时间内变得非常地非理性。在那个人能够阻止我之前，我把那瓶药喝了下去。没用几秒钟时间，我就显然变得疯狂了。我一边在房间里踉踉跄跄地走着，一边对那个人说："请便。我爱我的孩子。所以请你杀了他们。"那个人试图通过折磨我来得到那些黄金。我喊着说："疼极了。所以请你继续。"

鉴于我所处的状态，这个人现在有些无可奈何。他在设法让我打开保险箱方面，一筹莫展。威胁和折磨不能迫使一个陷入如此非理性状态的人屈服。这个人只有逃跑，希望躲过警察。而且既然我处于这样一种状态，他不大相信我还能记下他的汽车牌照，从而少了一些杀我的理由。

当我处于这样的状态时，我将以一些非理性的方式行动。所以存在这样一种风险，就是在警察到来之前我可能伤害我自己或者我的孩子。但是由于我没有枪，这个风险是小的。而造成我自己非理性就是降低这个人杀死我们大家这个大风险的最好方式。

按照任何一个令人信服的有关合理性的理论，在这种情况下使自己在一段时间内成为非理性的这一点本身对我来说会是合乎理性的。这回答了我上文所提的问题。S可能告诉我们要使自己乐于以S宣称非理性的那些方式去行动。这绝不是对S的什么诘难。正如刚才的事例所表明的，一个可以接受的有关合理性的理论能够告诉我们要去做那种以它自身标准衡量属于非理性的事情。

下面考虑一个一般性的主张，这个主张有时表达为：

> （G1） 如果某个动机既是（a）对某个人来说使他自己抱有这个动机是合乎理性的，又是（b）对他来说使他自己失去这个动机是非理性的，那么（c）对这个人来说，按照这个动机行动不可能是非理性的。

在刚刚表述的那个事例中，当这个人仍然在我家的时候，对我来说使自己不再非理性会是非理性的。在这段时间内，我具有这样一组动机，就其而言（a）和（b）皆真。但是（c）为假。在这段时间内，我的行动是非理性的。我们因而应当拒斥（G1）。我们能够代之以这样的主张，既然对我而言使自己像这个样子是合理的，这就是一个合理的非理性事例。

6.何以自利论蕴含着我们不可能避免非理性地行动

回想一下凯特的情形，她接受有关自利的快乐论。我们有可能采纳某个别的理论。但是按照这些别的理论，会有一些事例在一些相关的方面类似凯特的事例。而且那些得出的主张可加以重述以便涵盖这些事例。

凯特最强烈的愿望是把自己的著作尽可能地写好，对她来说这是最好的。但是由于此缘故，她经常非常努力地工作，使自己在一段时间内精疲力竭和相当压抑。由于凯特是一位快乐主义者，所以她相信，当她以这种方式行动时是在做对她来说更糟的事情。因为她还接受自利论，凯特相信，在这些情形中她是在非理性地行动。此外这些

非理性的行动还是完全自愿的。她像她所做的那样行动是因为，尽管她在意自己的利益，但这并不是她的最强烈愿望。她有一个甚至更加强烈的愿望，那就是她的著作尽可能地好。如果这个愿望减弱的话，对她来说会更糟。她是在按照这样一组动机行动，根据S，对她来说造成自己失去这些动机会是非理性的。

有人或许主张说，因为凯特是在按照这样一些动机行动，她不可能是在非理性地行动。但是这个主张假定有（G1），而（G1）这一主张被我所称的"谢林对武装抢劫的回应"事例表明为假。

如果我们分享凯特有关她是在以一种完全自愿的方式非理性地行动这一信念，那么，我们可能主张她是非理性的。但是凯特能够予以否认。既然她相信S，她能够主张说："当我做我相信对我更糟的事情的时候，我的行动是非理性的。但是因为我是在按照这样的动机行动的，而对我来说使自己失去这些动机会是非理性的，所以我并非是非理性的。更确切些说，我是合理性地非理性的。"

她还可以补充说："在按照我想把我的书写得更好这样的愿望行动方面，我是在做对我来说将更糟的事情。这是一个以利己标准衡量的糟糕后果。但它是作为最好的可能后果组之一的一组后果的部分。我尽管有时受苦，但是因为这是我的最强烈愿望，所以我也受益。而且那些获益大于那些损失。我有时非理性地行动，做我自知对我将更糟的事情，这只是我倘若要得到这些更大利益的话所不得不付出的代价。这是值得付出的代价。"

它可能遇到这样的诘难："你并非一定要付出这个代价。你可以工作少卖力一点。你可以做对你会更好的事情。你并非被迫去做你相信是非理性的事情。"

她会这样回答："的确如此。我可以工作少卖力一点。但是，只有在我要使我的著作写得更好的愿望大大减弱的时候，我才会这么做。而且这总体上来说对我会更糟。它会使我的工作索然无味。我怎么会导致产生这么一点，即，在未来的这样的一些情况下我将不是自由地

选择去做我认为是非理性的事情？只有通过一种对我会更糟的方式改变我的愿望，我才会导致这一点的产生。正是在这个意义上，我不付出较小的代价就不能够得到那些更大的利益。若非有时自由地选择以那些将对我更糟的方式行动，我就不能够抱有对我来说最好的那些愿望。这就是为什么，当我以这些方式非理性地行动的时候，我无需把我自己当作非理性的。”

这个回答假定了关于那些自愿行动的一个观点：心理决定论。据此观点，我们的行动总是由我们的愿望、信念和其他一些意向导致的。鉴于我们的实际愿望和意向，我们以不同方式行动并不是必然地可能的。它可能遭遇这样的诘难：“如果凯特与众不同地行动并不是必然地可能的，那么她就不应当相信，要理性地行动，她就应该与众不同地行动。我们只应该做我们能够做的。”

当讨论到道德上我们应该做什么的时候，会出现一个类似的诘难。如果凯特一并回答这两个诘难会节省一些文字。她会回答说：“在应该蕴含能够这个学说中，‘能够’的意思与心理决定论相容。当我的行动是非理性的或者失当的时候，我本应该以某种其他方式行动。按照该学说，只有当我本来能够这么做的时候，我才本该以其他方式行动。如果我本来就不能够以其他方式行动，就不能主张说，这是我本来应该做的。主张（1）：我本来就不能够以其他方式行动，并非主张（2）：鉴于我们的实际愿望和意向以这种方式行动本来就是不可能的。主张毋宁是（3）：即便我的愿望和意向有所不同，以这个方式行动本来也是不可能的。无论我们的愿望和意向可能是什么，以这种方式行动本来就是不可能的。如果主张（1）是主张（2）的话，那么，决定论者本会不得不得出任何人都不可能错误地或者非理性地行动的结论。但是他们能够具有正当性地拒斥这个结论。他们能够坚持认为主张（1）是主张（3）。”

凯特会补充说：“我并不主张自由意志与决定论兼容。自由意志的‘能够’的意思不同于那个蕴含能够学说中的‘能够’的意思。绝大多数

认为自由意志与决定论不相容的决定论者都认为这两个意思是不同的。这就是为什么，尽管这些决定论者并不认为任何人该当惩罚，但是他们继续认为错误地或者非理性地行动仍然是可能的。”

凯特假定心理决定论可能是错误的。我早先做出过这样的主张，我们有关合理性的信念可能影响我们的行动，因为我们可能想要合理地去行动。它可能会遇到这样的诘难：

> 这错误地描述了这些信念如何影响到我们的行动。我们并非通过援引某人这么做的愿望来解释为什么某人合理地行动。每当有人合理地行动的时候，他是想要这么做这一点可能真是微不足道的。但是他像他所做的那样行动了，是因为他具有一个信念，而不是一个信念和一个愿望。他像他所做的那样行动了，只是因为他认为他有一个这么做的理由。而且无论他的愿望和意向如何，对他来说合理地行动经常是因果地可能的。[4]

要注意，这个诘难者不可能主张，无论一个人的愿望和意向如何，对他而言合乎理性地行动总是可能的。即便他否定决定论，这个诘难者也不能够主张，在我们的行动和我们的意向之间没有什么联系。

这个诘难者还必须承认，我们的愿望和意向可能使我们更难以做我们相信是合理的事情。假设我在遭受极度的干渴之苦时，获得了一杯冰水。而且假设我有一个慢慢地喝下这杯水的理由，因为这会增加我的享受。我也有一个不溅出这杯水的理由。按照这第二个理由去做比在我极度干渴的情况下慢慢地喝下这杯水要更容易得多。

如果诘难者的观点为真，那么凯特的回应就必须予以修正。她或许说：“如果我最强烈的愿望是避免做我认为是非理性的事情的话，那么这对我会更糟。我的最强烈的愿望是我的著作尽可能地好，这对我会更好。既然这是我最强烈的愿望，那么我有时会做我认为是非理性的事情。我之所以以这种方式行动，是因为我要使我的著作尽可能地好的愿望比我不要非理性地行动的愿望强烈得多。你主张说我会时常避免以这种方式行动。凭借意志活动，我会时常避免做我最想做的事

情。如果我会避免以这种方式行动的话，我就能够主张我决不是非理性的。但是，鉴于我要使我的著作更好的愿望很强烈，对我来说避免以这种方式行动会是非常难的。而且对我来说，改变自己的愿望以便我自己更加容易地避免以这种方式行动，会是非理性的。鉴于这些事实，我只是在一个非常弱的意义上说是非理性的。”

凯特或许补充说：“我既有那些以利己标准衡量的最好的一组可能的动机中的一个，又永不做我所认为的非理性的事情，这是不可能的。说这不可能是在这样一个相关的意义上：无论我的愿望和意向是什么，这都是不可能的。如果我从不克己，我的日常行动就从来不会是非理性的。但是在造成我自己成为或任由自己保持永不克己时，我本会非理性地行动。如果代之以我使自己具有那些最好的一组可能动机中的一个，我将有时做我认为是非理性的事情。如果我并不具有某个从不克己的人的那种意向，那么，我总是如某个具有这一意向的人那样行动，就是不可能的。既然这是不可能的，而且对我来说使自己永不克己会是非理性的，那么我不能够因为有时做我认为是非理性的事情而受到批评。”

现在可能有人说，就像凯特所表述的，S缺乏任何一种理论的本质特征中的一个。可能有这样的诘难：“没有什么理论能够要求不可能的事。既然凯特不能够总是避免做S宣称为非理性的事情，她就不能够做S主张她应该去做的事情。因此，我们应当拒斥S。如同前述，应该蕴含能够。”

即便我们否定决定论，这个诘难仍然适用。正如我已经主张的，我们必须承认，既然凯特并不具有某个永不克己的人的那种意向，她就不能够总是像这样的一个人那样去行动。

凯特不能够总是避免做S宣称为非理性的事情这一点是对S的一个好的诘难吗？回想一下“谢林对武装抢劫的回应”。在这个事例中，按照任何有关合理性的令人信服的理论，对我来说不使自己非常地非理性会是非理性的。但是，如果我果真使自己非常地非理性，我就不能

够避免非理性地行动。按照这两种选择余地，至少我的行动中的一个会是非理性的。因而在这个事例中我不能够避免非理性地行动就是真的。既然能够有这样的一些情形，那么一个可以接受的理论就会意味着我们不能够避免非理性地行动。S具有这一点蕴含不是对S的什么诘难。

我们可能认为这些主张并不完全解答这个诘难。后面还将出现针对某些道德理论的一个类似诘难。为节省文字起见，我把这些诘难一并放在第15节中予以讨论。

我现在将总括一下我的其他一些结论。就许多人的情况而言，而且恐怕就绝大多数人的情况而言，自利论是间接地自败的。就这些人中的每一个人而言，如果他从不克己的话——永不准备做他认为对他会更糟的事情，反而对他会更糟，此为真。如果他具有的是某一组其他动机的话，会更好一些。我业已提出主张，这样的一些情形并不提供对S的一个诘难。既然S并不告诉这些人要永不克己，而是告诉他们，如果他们能够的话，不要永不克己，那么S以其自身的标准衡量不是失败的。这些情形也不提供对S的一个独立诘难。

尽管他们并没驳倒S，但是对那些接受S的人们来说这些情形却具有非常大的重要性。在这些情形中，S所涵盖的不仅是日常的行动，而且涵盖那些导致我们动机出现变化的行动。根据S，使我们自己具有或保持那些以利己的标准衡量最好的数组动机中的一组会是合理的。如果我们相信我们可以以这两种方式中的任何一种方式行动，不去这么做就会是非理性的。就绝大多数人的情况而言，最好的数组动机中的任何一组都会使他们有时以一种完全自愿的方式去做他们明知对他们将更糟的事情。如果这些人相信S，他们将认为这些行动是非理性的。但是他们无需认为他们自己是非理性的。这是因为，根据S，对他们来说，他们改变动机以便自己不会再以这种方式非理性地行动，反而会是非理性的。他们将在一定程度上痛惜这些非理性行动

的那些后果。但是，对于这些行动的非理性他们能够悠然自得地予以对待。这是合理的非理性。

这些主张可能遇到这样的诘难，就是它们错误地假定心理决定论。无论他们的愿望和意向如何，对这些人们而言做他们认为是合理的事情有时是可能的。如果这个诘难成立，这些主张就需要修正。当这些人做他们认为是非理性的事情的时候，他们就不能够主张他们决不是非理性的。但是他们能够主张，鉴于他们的实际动机，对他们来说避免以这种方式行动是非常艰难的。而且按照他们的理论，对他们来说改变自己的动机以便更容易避免以这种方式行动会是非理性的。因而他们能够主张，他们只是在非常弱的意义上是非理性的。一经解释如何能够对这些主张加以修正之后，只要下文不相干，我就不再提及这一诘难。对类似的任何主张作出所需的修正不应是件难事。

7.当与道德相冲突时对自利论予以拒斥的一个论证

有人论证过，自利论也许告诉我们要相信的不是它自己而是某个别的理论。这显然是可能的。根据S，如果对自己会更好的话，我们每一个人使自己相信某个别的理论会是合理的。

我已经提到过这可能为真的一种方式。对我们来说做我们认为非理性的事情或许是不可能的。自利论会告诉我们，在我一直在讨论的那些例子中，要努力相信一个与之不同的理论。这一点在其他方面也或许为真。让我们回到，比方说，遵守诺言这个问题。

达成一种双向协议在实践上是极为重要的。在这些协议中，某个群体中的每一个人都做出了有条件的承诺。每一个人承诺以一种特定的方式行动，条件是所有其他人都承诺以某些特定的方式行动。下面两者都能够为真：（1）如果所有的人而不是没有人遵守诺言，这对这些人中的每一个人都将更好，而且（2）无论其他人做什么，对每一个遵守诺言的人而言将更糟。如果他遵守自己的诺言每一个人所失去的要少于如果所有其他人都遵守他们的诺言他所得到的。这就是何

以（1）和（2）皆真。这样的一些协议是双向有利的，尽管要求克己。

如果人们知道我从不克己，我将被排除在这样一些协议之外。其他人知道，不能相信我会遵守诺言。已经有这样的主张：既然这是真的，如果我不再永不克己而且变得值得信赖的话，那么对我会更好一些。[5]

该主张忽视了一个可能性。如果我显得值得信赖但实际上仍然永不克己，这可能对我是最好的。既然我显得值得信赖，其他人就会允许我加入这些互利的协议。因为我实在从不克己，所以每当违背诺言对我更好的时候，我就会坐收不守诺言之利。既然对我来说显得值得信赖更好，对我来说遵守诺言以保持这种表面现象通常更好。但是有一些诺言我能够秘而不宣地予以违背。而且我从违背某些诺言中之所得可能大于不再显得值得信赖之所失。

不过，假设我是坦率的，无法令人信服地说谎。很多人的情况都是如此。如果我们开发出价格低廉却十分准确的测谎仪测试，此种情况的范围或许会变得更大。我们姑且假设这种情况已经发生，所以我们都是坦率的——不再能够欺骗他人。既然我们在某种程度上是坦率的，所以我的结论可能适用于我们的实际状况。但是，假定所有的直接欺骗已经变得不可能，将简化该论证。这样的一个论证表明什么，值得一看。因此，姑且承认这个假定，应当有助于这个论证。

如果我们都是坦率的，那么我们每一个人倘若变得值得信赖——即使当他相信这样做对他更糟的时候，他也可靠地倾向于遵守诺言——对他自己会更好。所以，根据S，我们每一个人都使自己值得信赖会是合理的。

下面假设，要想变得值得信赖，我们得改变有关合理性的信念。我们会不得不使自己相信，对我们每一个人来说守诺是合理的，即便当他知道这对他更糟的时候也要如此。我将在更后面的某部分文章中表述这个假定或许为真的两种方式。

当我们据以改变信念的理由只是这个改变将符合我们的利益的时候，这个改变是艰难的。我们会不得不使用某种形式的自欺。比方说，假设我获悉我得了致命的疾病。既然我想要相信自己是健康的，那么我付费给催眠师让他带给我这种信念。如果我记得是如何得到这个信念的话，我就不会保持这个信念。如果我记得这一点的话，我会知道这个信念是假的。我们有关合理性的信念情况同样会如此。如果我们付费给催眠师让他改变我们的这些信念，因为这对我们将变得更好，那么，那些催眠师必须使我们忘记我们为什么具有我们的这些新信念。

按照上述有关假定，S会告诉我们要改变我们的信念。S会告诉我们要相信S的一个修正形式，而不是它自己。按照这个修正后的理论，我们每一个人去做认为对自己更糟的事情是非理性的，除非他是在遵守诺言。

如果S告诉我们去相信这个修正后的理论的话，这会是对S的一个诘难吗？它会表明遵守这样一些承诺是合理的吗？我们必须清楚地关注这个问题。我们相信这样一点可能是恰当的，就是，遵守我们的诺言是合理的，即使当我们知道这对我们将更糟的时候。我要问的是："如果S本身告诉我们要使我们自己具有这个信念的话，这个信念是得到了支持的吗？"

一些人给出的答案是肯定的。他们论证说，如果S告诉我们要使自己具有这个信念，这就表明这个信念是得到正当性证明的。而且他们把这个论证应用于许多别的种类的行动，他们认为这些种类的行动像遵守诺言一样是道德上所要求的。如果这个论证取得了成功的话，它本来会有极大的重要性。它会表明，在许多种类的情形中，合乎道德地行动是合理的，即便是当我们相信这对我们将更糟的时候。道德上的理由会表明比自我利益所提供的那些理由更强一些。许多作者尝试过证明这个结论的正当性，但没有成功。如果这个结论能以刚刚提

及的那种方式证明其正当性的话，就会解决西季威克所称的"伦理学的最为深刻的难题"了。[6]

8.这个论证何以失败

对这个论证有一个简明的诘难。该论证诉诸的是这样一个事实，那就是，S会告诉我们要使自己相信，即便当我们知道遵守诺言对我们将更糟的时候，遵守诺言也是合理的。称这个信念为B。B与S不相容，因为S主张，遵守这样的承诺是非理性的。S要么是有关合理性的真的理论，要么不是。如果S为真，那么B必然假，因为它与S不相容。如果S不为真，B或许为真，但是S并不能够支持B，因为一个不为真的理论不能够支持任何结论。简而言之，如果S为真，B必假，而且如果S不为真，那么它不能够支持B。B要么是假的，要么是得不到支持的。是故，即使S告诉我们要力求相信B，这个事实也不能够支持B。

我们可能认为，一个有关合理性的理论不能够为真，但充其量能够是最好的或者是得到最好的正当性证明的理论。刚刚给出的诘难可以从这样的角度予以重新陈述。存在两种可能性。如果S是最好的理论，我们就应当拒斥B，因为它与S并不相容。如果S并不是最好的理论，我们就应当拒斥S。B不能够得到一个我们应当予以拒斥的理论的支持。这两种可能性中没有哪一个给予B任何支持。[7]

这个诘难在我看来是强劲的。但是我知道它并未使一些人信服。因而我将再给出两个诘难。这些诘难也将支持某些更加广泛的结论。

我将首先区分威胁和警告。当我说除非你做Y否则我将做X的时候，倘若我做X对你而不是对我会更糟，称之为警告；倘若我做X对我们两者都会更糟，则称之为威胁。如果我总会实施我的威胁，则称我为实施威胁者。

假设，除了是一个实施威胁者之外，某个人永不克己。即使这样的人知道实施他的威胁对他而言更糟，他也会实施。但是，如果他认

为做出威胁对他更糟的话，他就不会做出威胁。这是因为，除了是一个实施威胁者之外，这个人从不克己。他从不做他所认为的对他将更糟的事情，除非他是在实施某个威胁。这个例外并不涵盖做出威胁。

假设我们都是既坦率，又永不克己。如果此点为真的话，倘若我使自己成为一个实施威胁者，并且随后向别的每个人宣布我的意向的这一变化，对我来说会更好。既然我是坦率的，所以每一个人都会相信我的威胁。而且被确信的威胁具有许多用处。我的有些威胁会是防卫性的，意在防范别人侵犯到我。我或许把自己局限于防卫性的威胁。但是，以其他方式利用我的那些众所周知的意向会有相当的诱惑力。假设我们之间分享某种合作的利益。而且假设，没有我的合作，就不会有进一步的利益。我或许说，除非我得到最大的份额，否则我将不予以合作。如果其他一些人知道我是一个实施威胁者，而且他们永不克己，他们将给我最大的份额。不这么做对他们会更糟。

其他一些实施威胁者或许以更糟的方式行动。他们会把我们贬为奴隶。他们会威胁说，除非我们成为他们的奴隶，否则他们会与我们同归于尽。我们知道这些人会实施他们的威胁。我们因而会知道，只有成为他们的奴隶才能够避免毁灭。

如果我们都是坦率的，对实施威胁者的回应就是变成不顾威胁者。这样的人总是不顾威胁，即使当他知道这样做对他更糟时也是如此。一个实施威胁者不会威胁到一个坦率的不顾威胁者。他会知道，如果他进行威胁的话，他的威胁会不被理睬，而且他一旦实施威胁，对他而言，情况只会变得更糟。

如果我们都是既坦率，又从不克己，我们意向中的什么变化会对我们每个人更好一些？我把这个问题放到"附录A"中回答，因为这个回答的有些方面与我们现在正在讨论的问题并不相干。相关的是这样一点。如果我们都是坦率的，对我们每一个人来说成为一个值得信赖的不顾威胁者很可能会更好一些。这两个变化会包含某些风险；但是那些很可能有的利益会大大超过这些风险。值得信赖会有什么好处？就

是我们不会被排斥在那些需要克己的互利的协议之外。成为不顾威胁者会有什么好处？就是我们不会成为那些实施威胁者们的奴隶。

接下来我们可以假定我们不会成为值得信赖的不顾威胁者，除非我们改变了我们有关合理性的信念。那些值得信赖的人们即使在知道将对他们更糟的时候也遵守他们的诺言。我们能够假定，除非我们相信遵守这样的诺言是合理的，否则我们就不会具有以这种方式行动的意向。而且我们可以假定，除非我们被人们认识到具有这个信念，否则其他人不会相信我们会遵守这些诺言。按照这些假定，S告诉我们要使我们自己具有这个信念。类似的说法适用于成为不顾威胁者。我们可以假定，除非我们相信不顾威胁总是合理的，否则我们就不会成为不顾威胁者。而且我们可以假定，除非我们具有这个信念，否则别人就不会信服我们是不顾威胁者。按照这些假定，S告诉我们要使自己具有这个信念。这些结论能够结合起来。S告诉我们要使自己相信，在做我们相信将对我们更糟的事情时总是非理性的，除非我们是在遵守诺言或是在不顾威胁。

这个事实支持这些信念吗？根据S，对我们每一个人来说，以下这点是合理的，使自己相信即使在自知将对自己更糟时也不顾威胁是合理的。这表明这个信念是正确的吗？它表明不顾这威胁是合理的吗？

举个例子将对我们有所帮助。考虑一下

我的奴隶状态。你和我分享一个荒岛。我们两个人都是坦率的，而且从不克己。现在你的意向出现一种变化，成为一个实施威胁者。而且你有一枚能够炸飞这个岛屿的炸弹。通过不断地威胁要引爆这枚炸弹，你迫使我为你艰苦地工作。对你的权力的惟一限制是，你必须让我值得活下去。如果连这一点都达不到的话，你的威胁对我来说就不再有效。

我如何才能够终止我的奴隶状态？杀死你不行，因为除非你定期拨动某些密码，否则，你的炸弹会自动爆炸。但是假设我会使自己成

为一个坦率的不顾威胁者。愚蠢透顶地，你没有威胁说你会不顾我的意向中的这一变化。这样，这个改变会终结我的奴隶状态。

对我来说作出这个变化会是合理的吗？毕竟存在着你可能作出某个新威胁的风险。但是，既然这样做明显会对你更糟，这个风险会是小的。而且通过甘冒这个小风险，我几乎会肯定地得到更大的益处。我几乎肯定会终结我的奴隶状态。鉴于我的奴隶状态的悲惨境地，对我来说，根据S，使自己成为一个不顾威胁者是合理的。而且鉴于我们的那些别的假定，对我来说使自己相信不顾威胁总是合理的这一点会是合理的。尽管我不能够完全肯定这会对我更好，但是那巨大的、几乎可以肯定的利益会超过那个小风险。（在相同的意义上，某个人成为值得信赖的话会对他更好这一点永远不会是完全肯定的。此处也是，所有能为真的就是：可能有的那些利益超过了那些风险。）

假定我现在作出了这些改变。我已经成为一个坦率的不顾威胁者，而且已经使自己相信不顾威胁总是合理的。根据S，对我来说使自己具有这个信念是合理的。这表明这个信念是正确的吗？

让我们继续那个故事。

> 我如何终结我的奴隶状态。我们两个都运气不好。有一段时间，你忘了我已经成了一个不顾威胁者。为了达到某个微不足道的目的——诸如得到我刚采摘的椰子——你重复你通常的威胁。你说，除非我给你那个椰子，否则，你会把我们两个炸上天。我知道，如果我拒绝，这对我将肯定更糟。我知道你确实是一个实施威胁者，即便知道这对你将更糟也会把威胁付诸实现。但是，像你一样，我现在不信那个纯粹的自利论。我现在相信，即便我知道这对我将更糟，不顾威胁也是合理的。我按照我的信念行动。正如我所预见到的，你把我们两个人都炸飞了。

我的行动是合理的吗？不是。如前所述，我们或许得出结论，既然我按照如下信念行动：对我来说，承认我不是非理性的，这一点是合理的。更准确地说，我是合理地非理性的。但是我所在做的却不是合理的。当我知道如果我做的话，这对我将是灾难性的，而且对什么人都没有好处，那么不顾某个威胁就是非理性的。S在此告诉我们，

使自己相信即使在自知将对自己更糟的时候也不顾威胁是合理的，这一点是合理的。但是这并不表明这个信念是正确的。并不表明在这样一个情形中不顾威胁是合理的。

我们能够得出一个更宽泛的结论。这个情形表明我们应当拒斥

（G2） 如果某个人使自己相信，对他而言以某种方式采取行动是合理的这一点是合理的，那么，对他而言以这种方式行动就是合理的。

现在回到B，即回到这样一个信念，即使当我们知道这对我们将更糟时还遵守我们的诺言是合理的。按照上述那些假定，S蕴含着对我们来说使我们自己相信B是合理的。有些人主张，这个事实支持B，表明遵守这样一些诺言是合理的。但是这个主张好像假定（G2），而这是我们刚刚拒斥了的。

对于这些人所主张的还有一个诘难。即使S告诉我们力求相信B，S也蕴含着B是假的。所以，如果B为真，S必假。既然这些人们相信B，他们应当相信S是假的。那么他们的主张会假定

（G3） 如果某个有关合理性的假理论告诉我们使我们自己具有一个特定的信念，这表明这个信念是真的。

但是我们显然应当拒斥（G3）。如果某个假理论告诉我们要使自己相信地球是方的，这不会表明地球就是如此。

S告诉我们要力求相信不顾威胁是合理的，即使我们知道，这么做对我们将更糟。正如我所举的例子所表明的，这并不支持这个信念。因此，我们应当就遵守诺言持有相同的主张。可能还有一些别的根据使我们可以据以相信遵守我们的诺言是合理的，即使我们知道这样做对我们将更糟。但是这并不会被这样一个事实——自利论本身告诉我们使自己相信它是合理的——表明为合理的。已经有这样的论证，就是通过诉诸这样一些事实，我们能够解决一个古老的难题：我

们能够表明，当它与自我利益相冲突的时候，道德提供更强的行动理由。这个论证站不住脚。它或许充其量表明某种东西更少。在一个我们都坦率的世界中——不能够互相欺骗——在有关合理性的方面自欺或许是合理的。[8]

9.何以自利论可能是自谦的

如果S告诉我们相信某个别的理论，它也不会支持这个别的理论。但是它会是对S的一个诘难吗？再重申一遍，以其自身标准衡量S不会是自败的。S是一个有关实践上的合理性而不是理论上的合理性的理论。S可能告诉我们，使我们自己抱有某个错谬的信念。如果对我们来说抱有错谬的信念会更好一些的话，抱有真的信念，即便是关于合理性的，也不会是S给我们设定的那个终极目标的部分。

上面所给出的那些论证或许可以得到强化和扩展。如果正如我所假设的，测谎技术使我们大家都坦率的话，这会更容易一些。如果我们永远不能够彼此欺骗的话，或许有一个论证表明，根据S，每一个人使自己不相信S会是合理的。

假设这是真的。假设S告诉每一个人要使自己相信某个别的理论。那么S会是自谦的。姑且假设我们都相信S，但是我们也可能改变我们的信念，那么S就会自己隐退。它就会变成一个没有人相信的理论。但是自谦并非是自败。它不是有待人们相信的一个理论的目标。如果我们对理论加以人格化，而且佯装它们有目标，那么一个理论的目标不是有待人们相信，而是有待证明其为真，或证明其为最好的理论。一个理论是自谦的并不表明它不是最好的理论。

如果我们相信S对我们来说会更糟时，S会是自谦的。但是S无需告诉我们要相信它。当如果我们相信某个别的理论会对我们更好时，S会告诉我们力求相信这个理论。如果我们成功地做到了S所告诉我们的，这对我们再一次地会更好一些。尽管S会自己隐退，致使没有什么人相信它，以自身的标准衡量它仍然不会是失败的。这一点仍然会

是真的，因为我们每一个人都已经遵循S——做了S所告诉我们要做的——每一个人因而已经使结果对自己来说更好了一些。

尽管以其自身标准衡量S不会是失败的，但是或许会有人主张说，一个可以接受的理论不可能是自谦的。我对这个主张予以否定。这个主张可能看起来令人信服，但是一经分析便显出糟糕的理由。想要一个关于合理性的最好的理论不是自谦的，这会是自然的。如果最好的理论是自谦的，告诉我们相信某个别的理论，有关合理性的真理就会令人沮丧地出现复杂化。希望真理更加简明一些：最好的理论会告诉我们相信它，这一点是自然的。但是能让它不止是一个希望吗？我们能够假定真理必定更加简明吗？我们不能。

10.何以后果论是间接自败的

我的绝大多数主张，经过小小的改动，就会涵盖一组道德理论。这些就是后果论（或用缩写C表示）的那些不同版本。C的中心主张是

（C1） 存在一个终极道德目标：结果尽可能地好。

C理论适用于一切事情。当应用于行动的时候，C主张下述两点：

（C2） 我们每一个人所应当做的是任何产生最好结果的事，而且
（C3） 如果某个人所做的是他相信将使结果更糟的事，他就是在错误地行动。

我在我们最有理由去做的事情与我们相信或者应该相信在既定情况下去做会是合理的事情之间作了区分。我们现在必须区分什么是客观的对与错和什么是主观的对与错。这个区分与道德理论是否能够客观上为真毫无关系。该区分是在下述两点已知的情况下某个理论所蕴含的东西之间的区分：（i）某个人所做的或本可以做的后果或本来有的后果，和（ii）这个人所相信的或应该相信的这些后果。

提及一个类似的区分可能会有所帮助。客观上正确的医治是事实上对病人最有好处的那个医治。主观上正确的医治是在医学证据已知的情况下医生最合理地开出处方的医治。正如这个例子所表明的，认识到的最好的东西是客观上正确的东西。道德理论的中心部分回答的就是这个问题。我们有两个理由需要说明主观上的正确性。我们通常并不知道我们行动的后果会是什么。而且我们应该为做了主观上错的事情而受到责难。即使这些行动客观上是对的，我们也应该为这样的一些行动而受到责难。即使一位医生所做的非常可能要了病人命的事情在事实上救了病人一命，他也应该为他的行动而受到责难。

在下文的绝大多数篇幅中我将在客观意义上使用对、该、好和坏。但是错将通常意指主观上的错或该受责备。我意指的意思在语境中常常是显而易见的。这样，就上面给出的主张而言，显然（C2）是关于什么是我们客观上该做的，而（C3）是关于什么是主观上错的。

为了涵盖那些带有风险的情形，C主张

（C4） 我们主观上应该做的是那些其结果具有所预期的最大的善的行动。

计算一个行动的结果的所预期的善的方法是，每一可能的善的后果的价值乘以该行动将产生该后果的几率。每一个可能的糟糕后果的负面价值也是如此计算。结果的所预期的善是这些价值减去这些负值后的总量。假设，例如，如果我往西行的话，我有1／4的几率救100个人的命，有3／4的几率救20个人的命。我的西行所预期的善，以所救的人数衡量，是100×1／4＋20×3／4，或25＋15，或40。下面假设我往东行，我将肯定救30个人的命。我东行所预期的善是30×1，或者30。根据（C4）我应该西行，因为所预期的得救人数会更大。

后果论不仅涵盖行动和结果，而且涵盖愿望、意向、信念、情绪、眼睛的颜色、气候以及其他任何东西。更确切些说，C涵盖任何会使结果更好或者更糟的事情。根据C，最好可能的气候是会使结果最好的气候。我将再次用“动机”来涵盖愿望和意向这两者。C主张

（C5） 最好的可能动机是那些如果我们具有它们结果就将为真的最好的动机。

如前所述，“可能”意指“因果上可能”。有许多不同的动机组在这样一个意义上会是最好的：没有什么别的这样的可能动机组，如果我们具有这个动机组的话，结果会真的更好。我已经表述过我们能够借以改变我们的动机的一些方式。（C2）蕴含着我们应该力求造成我们自己具有或保持任何一组最好的可能动机组。更宽泛一些说，我们应该以任何一种会使结果更好的方式来既改变我们自己又改变其他任何东西。如果我们相信我们本来能造成这样的改变，那么（C3）蕴含着未能这样做会是错的。[9]

要应用C，我们必须问是什么造成结果更好或者更糟。最简洁的答案是由功利主义给出的。这个理论把C与下面这样一个主张结合起来：最好的结果是那个带给人们减去负担后的最大净利益总量的结果，或者按照这个主张的快乐主义版本来说，是那个带给人们减去悲惨之后最大净幸福总量的结果。

C的版本还有很多。这些版本能够是一些多元主义理论，诉诸数个不同的有关是什么造成结果更好或者更糟的原则。因此，C的一个版本既诉诸那个功利主义主张，又诉诸平等原则。这个原则主张，如果某些人不是因为自身的过错而境况比别人更糟，那么这是很糟糕的。按照C的这个版本，一个结果的善既有赖于利益的净总量会有多大，又有赖于利益和负担在不同的人们之间分配的平等程度如何。尽管两个结果中的一个牵涉到一个较小一些的利益总量，但两个结果中的这一个结果或许更好一些，因为这些利益会得到更为平等的分享。

一位后果论者会诉诸许多别的原则。根据三个这样的原则，倘若人们受到欺骗、遭到强迫和背叛，是很糟糕的。而且这些原则中的一些可能本质上是指过去的事件。两个这样的原则就诉诸过去的资格和公正的应有赏罚。平等原则可能主张，人们应当得到平等的份额，不是在某些特定的时刻如此，而是在他的整个生活中如此。如果它提出这个主张的话，这个原则本质上指涉过去的事件。如果我们的道德理

论包含这样一些原则，我们就不只是关切狭义上的后果：我们行动之后所发生的结果。但是我们仍然能够在更广的意义上是后果论者。在这个广义上，我们的终极道德目标不是结果尽可能地好，而是历史尽可能好地进行。我下面所说的或许能够从此角度予以重新陈述。

我将用语词“后果论”和字母“C”指称所有这些不同的理论。就像有关自利的那些不同的理论一样，在C的这些不同版本之间作出决断至少也会占用一本书的篇幅。本书并不讨论这个决断。我只讨论这些不同的版本有什么共同之处。我的那些论证和结论会适用于所有的或几乎所有的此类具有可信度的理论。值得强调的是，如果一个后果论者诉诸我提到的所有原则，那么他的道德理论就非常不同于功利主义。既然这样一些理论很少被讨论到，所以可以轻易地忽略不计。

一些人已经想到，如果后果论诉诸许多不同的原则，那么它就不再是一个与众不同的理论，因为它能够被用来涵盖所有的道德理论。这是一个误解。C只是诉诸那些关乎什么使结果更好或更糟的原则。于是C或许主张，如果存在更多的欺骗或者强迫，这会更糟。那么C会给我们大家设定两个共同目标。我们应当力求使欺骗或强迫更少这一点属实。既然C给所有的行动者设定共同的道德目标，我将称C是对行动者中立的。

许多道德理论并不采取这样的形式。这些理论是相对于行动者的，给不同的行动者设定不同的目标。例如，能够这样主张，就是每一个人应当具有他不去强迫他人这一目标。照此观点，我强迫他人会是错的，即便这样做会使强迫更少一些这一点为真。这一主张用在欺骗或背叛他人上面也是同样。按照这些主张，每一个人的目标不应当是欺骗或背叛更少一些，而应当是他自己并不欺骗或背叛他人。这些主张不是后果论的。而且这些是我们绝大多数人都接受的那种主张。C能够诉诸有关欺骗和背叛的那些原则，但是它并不诉诸这些原则的熟悉形式。

我现在将表述某个理论T可能是自败的一种不同方式：

如果当几个人力图达到他们的T设目标，而达到这些目标却将会更糟这一点为真时，我们称T是间接地集体性地自败的。

按照它的所有不同版本或者大多数不同版本，这一点就C的情况来说可能为真。C蕴含着我们总是应当力求做任何会使结果尽可能好的事情。如果我们倾向于以这种方式行动，我们是纯粹的行善者。如果我们都是纯粹的行善者，这可能会使结果更糟。即使我们过去总是做（我们的行动所可能做出的）会使结果最好的事情，此点仍然可能为真。糟糕的后果可能不是来自我们的行动，而是来自我们的意向。

有许多方式，如果按这些方式我们都是纯粹的行善者的话，这可能具有糟糕的后果。其中的一个是对幸福总量的后果。按照C的任何一个令人信服的版本，幸福是那种使结果更好的东西中的一个巨大部分。我们的绝大多数幸福来自具有某些强烈的愿望，并且按此行动。这些包括爱某些他人所牵涉到的愿望，好好工作的愿望，以及我们不工作时据以行动的许多强烈愿望。为要成为纯粹的行善者，我们会不得不有悖于甚或压抑这些愿望中的绝大多数愿望。很有可能的是，这会大大减少幸福总量。这会使结果更糟，即便我们过去总是做（我们的行动所可能做出的）会使结果最好的事情。鉴于人们的实际情况，它或许并不使结果比实际上的更糟。但是它会使结果，比我们若不是纯粹的行善者，而是具有某些别的因果上可能的愿望和意向的话本会有的结果，更糟一些。[10]

还有别的几种方式，如果我们都是纯粹的行善者的话，会使结果更糟。其中的一个基于这样一个事实，就是当我们想要以某些方式行动的时候，我们将很有可能在有关我们的行动后果方面欺骗我们自己。我们将很有可能错误地认为，这些行动将产生最好的结果。例如，考虑一下杀死其他人的例子。如果我们想要某个人死，就轻易地相信，错误地相信，这会使结果更好一些。它因而使这样一个结果更好一些：我们强烈地倾向于不杀人，即便当我们相信这样做会使结果更好一些时也是如此。只有当我们相信凭借杀会使结果更加多得好的

时候，我们不杀人的意向才应当让路。类似的主张适用于欺骗、强迫以及别的数种行动。

11.以其自身标准衡量何以后果论并未失败

我将假定，在这些和其他一些方面，后果论（C）是间接地集体地自败的。如果我们都是纯粹的行善者，结果会比我们如果具有某些别的动机组会有的结果更糟。如果我们知道这一点，C告诉我们使我们自己成为或者保持纯粹的行善者会是错的。因为C作出这个主张，以其自身标准衡量它并不是失败的。C并非宣告自己不适用。

对C的这种辩护就像我对S的辩护。有一个差异值得指出。当某个人的下述情况为真时，S是间接地个别地自败的：如果某个人从不克己的话，这会比倘若他有某组别的愿望和意向的话更糟。这会是一个以S的标准衡量的糟糕后果。而且这个糟糕后果时常发生。有许多人，因为从不克己或很少克己，他们的生活正在变得更糟。如果我们中的某些人或所有的人是纯粹行善者的话，这会使结果比倘若我们具有某些别的动机会有的更糟，当此点为真时，C是间接地集体地自败的。这会是一个以C的标准衡量的糟糕后果。但是这个糟糕后果可能不发生。没有多少人是纯粹的行善者。因为这样的人为数不多，他们具有这个意向这一事实总体说来可能并不使结果更糟。

以S的标准衡量的那个糟糕后果时常发生。以C的标准衡量的那个糟糕后果可能不发生。但是这个差异并不影响到我对S和C的辩护。两个理论都告诉我们不要具有产生这些糟糕后果的意向。这就是为什么以他们自己的标准衡量S并不失败而且C不会失败的缘由。与这些糟糕后果事实上是否发生并不相关。

我对C的辩护假定，我们能够改变我们的意向。可能遇到诘难说："假设我们都是纯粹的行善者，因为我们相信C。而且假设我们不能改变我们的意向。这些意向会具有以C的标准衡量的糟糕后果，而且这些糟糕后果会是我们相信C的结果。所以C以其自身的标准衡量是

失败的。”对于我对S的辩护也存在过一个类似的诘难。我将在第18节中讨论这些诘难。

12.幻想的伦理学

我已经假定C是间接地集体地自败的。我已经假定，如果我们都是纯粹的行善者，结果会比我们倘若具有某些别的动机组的话更糟。如果这个主张为真，C告诉我们，我们应当力求具有这些别的动机组中的一组。

这个主张是否为真部分地是一个事实性的问题。我认为它很有可能是真的。但是我将不会在此尝试表明这一点。这个主张蕴含什么看起来更值得讨论。我还相信，即便我们信服后果论是最好的道德理论，我们绝大多数人也不会事实上成为纯粹的行善者。

由于麦基作出一个类似的假定，他称行为功利主义为“幻想的伦理学”。[11]像其他的几个作者一样，他假定，我们应当拒斥这样一种道德理论，倘若它在这样一个意义上不切实际地要求：即便我们都接受这个理论，我们绝大多数人事实上也很少会做这个理论所主张我们该做的。麦基认为，道德理论是某种我们发明出来的东西。如果是这样，那么主张一种可以接受的理论不能够不切实际地要求这一点就是令人信服的。但是按照其他几个有关道德本质的观点，这个主张并不令人信服。我们可能希望最好的理论不是不切实际地要求的。但是，按照这些观点，这也只能是一个希望。我们不能够假定它必定为真。

假设，我错误地假定C是间接地集体地自败的。即便这是错谬的，我们也能够令人信服地假定C是不切实际地要求的。即便我们都是纯粹的行善者不会使结果更糟，我们大家或我们中间的绝大多数成为纯粹的行善者也很有可能在因果上是不可能的。

尽管这是些相当不同的假定，但是它们具有同样的蕴含。如果我们成为纯粹的行善者在因果上是不可能的，那么C再次蕴含我们应该力求具有以后果论的标准衡量的最好的可能动机组中的一组。如果

（1）C要么是间接地自败或是不切实际地要求的，要么两者都是，而且（2）这些事实中没有哪一个会表明C不能够是最好的理论，那么这个蕴含因而是值得讨论的。尽管我还没有确信C是最好的理论，但是我既相信（1）又相信（2）。

13.集体性的后果论

值得把C与后果论的另一种形式作一个区分。正如迄今所陈述的，C是个体性的，而且关乎实际的后果。根据C，鉴于其他人实际上将做的，我们每一个人都应当力求做会使结果最好的事情。而且鉴于其他人所具有的那些实际的动机组，我们每一个人都应当力求具有那些可能的动机组中其后果最好的那一组动机。我们每一个人都应当问：“存在某个这样的别的动机组吗，它对我来说既是可能的，结果又会是更好的？”我们的回答有赖于关于其他人将拥有的那些动机组，我们所知道的或是我们能预测到的。

当我在1983年1月打出这些文字时我能够预测到什么？我知道我们绝大多数人将继续具有那些与我们现在所具有的那些动机非常像的动机。我们绝大多数人都将爱一些特定的他人，而且将具有绝大多数幸福所赖以获得的其他强烈愿望。既然我知道此点，C可能告诉我力求成为一个纯粹的行善者。这可能会使结果更好，即使倘若我们都是纯粹行善者的话会使结果更糟。如果绝大多数人都不是纯粹的行善者，而一些人是的话，这可能使结果更好。如果绝大多数人仍然保持现在的样子，那么，将有更多的苦难，更多的不平等，更多的其他绝大部分使结果糟糕的东西。这种苦难中的许多我本能够相当轻易地避免，而且我本能够以其他方式做许多使结果更好的事情。如果我避免密切的人际关系，而且使我的其他强烈愿望变得相对更弱一些，以致我能是一个纯粹的行善者，那么这从而可能使结果更好一些。

如果我幸运的话，对我来说变得像这个样子可能并不坏。我的生活将被剥夺掉绝大多数幸福来源。但幸福的一个来源是与人为善的信

念。这个信念可能带给我幸福，使我朴素的生活不仅在道德上是好的，而且对我来说也是好的生活。

我也可能不那么幸运。尽管我会接近于做一个纯粹的行善者，但这的确对我可能不会是一个好的生活。而且可能存在其他许多对我可能更好的生活。按照有关自利的那些令人信服的绝大多数理论都可能会这样。那么C对我的要求可能看来有些不公。剥夺掉自己绝大多数幸福来源的人为什么应当是我？更确切些说，我为什么应当属于为数不多的根据C应当力求做到这一点的人？如果所有的人都更多地使结果更好的话，难道不会更公平吗？

这令人想到后果论的一个形式，它既是集体的，又关乎理想的后果。按照此一理论，我们每一个人都应当力求具有这样的一些愿望和意向组中的一组，如果每一个人都具有这些组中的一组的话，结果会比每一个人具有别的一些组的话更好一些。这个陈述能够以数个方式予以诠释，而且在去除模棱两可性方面存在一些人所共知的困难。此外，这个理论的某些版本面临强烈的诘难。它们告诫我们所要忽略的是事实上会发生的事，而且那些忽略方式可能是灾难性的。但是集体性的后果论（缩写CC），具有更大的吸引力。我将留待更后面一点提出，一个更复杂的理论是如何可能保持CC的吸引人之处而又避免那些诘难的。

CC并非只是在有关我们的愿望和意向方面与C不同。这两个理论还在有关我们应该做什么的问题方面见解不一。考虑一下富人应当送给穷人到底多少的问题。对绝大多数后果论者而言，这个问题超越了国界。既然我知道其他绝大多数富人只会给出一点点，对我来说会难以否定：如果我几乎给出我的全部收入的话会更好。即便我给出了十分之九，但是所剩的那十分之一中的一些如果让非常贫穷的人享用的话，行善会更多。后果论于是告诉我，应该给出我的几乎所有的收入。

集体性的后果论的要求要小得多。它并不告诉我给出事实上会使结果最好的那个数量。它告诉我给出这样多的量，就是如果我们都给出这个量的话，结果会最好。更确切些说，它告诉我给出那个会使结果最好的、特定的国际所得税所要求的量。这个税会是递增性的，对越富有的人所要求的比例越大。但是它对每一个人的要求会比C（按照对他人的实际付出数量的可信预测）所作出的要求小得多。如果富有程度像我的那些人都只给出他们一半的收入，或者只给出他们四分之一的收入，或许是最好的。可能为真的是：如果我们都给出更多一些，这将破坏我们自己的经济，以致在将来我们只能给出少得多。可能为真的是，如果我们都给出更多一些的话，我们的馈赠将会大得让那些贫穷国度的经济无法吸收。

我所讨论的这个差异只有在所谓的部分依从理论中才出现。这是道德理论中的这样一个部分，它涵盖我们知道某些他人将不履行他们应该做的事的那些情形。C或许要求一些人给出他几乎所有的钱财，并且力求使他们自己成为纯粹的行善者。但是这可能只是因为绝大多数其他人不是在做C所主张的他们应该做的。他们不向穷人给出他们应该给出的数量。

在其部分依从理论中，C一直被认为是要求过高的。此主张不是那个有关C是不切实际地要求的主张。正如我已经说过的，我认为这不会是什么诘难。所被主张的是，在其部分依从理论中，C提出了不公的或者说不合理的要求。这个诘难可能并不适用于C的完全依从理论。如果我们都具有，根据C我们应该力求促使自己具有的那些可能的动机组中的一组的话，C的要求就会小得多。[12]

14.无过之错

尽管C是间接地自败的，但是它以其自身标准衡量并非是失败的。但是它可能面临其他诘难。比方说我在讨论S的时候所提出的那些诘难。假设我们全都相信C，而且全都具有属于以后果论的标准衡

量是最好的可能动机组中的动机。我业已主张，至少对我们绝大多数人而言这些动机组不会包括做一个纯粹的行善者。如果我们不是纯粹的行善者，那么我们有时将做我们所相信的将使结果更糟的事。根据C，那么我们将是在错误地行动。

在此举一个例子。绝大多数最好的可能动机组中会包括我们自己对子女的强烈的爱。假设克莱尔具有这样的一个动机。考虑一下

> 事例一。克莱尔要么会给她的孩子某种利益，要么给某个不幸的陌生人更多一些利益。因为她爱她的孩子，所以她把利益给了孩子而不是那个陌生人。

作为一个后果论者，克莱尔可能不仅看重孩子在道德上得益多少，而且看重他们是否得益于他们自己的父母。她可能认为父母的关爱内在地是或本身就是使结果更好的东西。即使如此，克莱尔可能相信，由于未能帮助那位陌生人，她是在使结果变得更糟。她可能因而相信她是在错误地行动。而且这个行动完全是自愿的。只要她想，她本来能够避免做她所相信是错的事。她未能如此做只是因为，与想要避免做错相比，她更想要有利于她的孩子。

如果某个人自由地做她相信是错的事情，那么她通常会面临严峻的道德批评。克莱尔认为自己应该承受这样的批评吗？作为一个后果论者，她会否认这一点。克莱尔的回应会像凯特主张她自己并非是非理性的时候的回应一样。克莱尔会说："我之所以做错，是因为我爱我的孩子。但是，造成自己失去这样的爱，对我来说会是错的。我使结果变得更坏导致了一个坏的结果。但这是总体上来说最好的可能的后果组中的一个后果组的部分。因此，对我来说，改变我的动机以便未来不会再以这种方式做错，这会是错的。既然如此，当我的确以这种方式做错时，我无需把我自己当作道德上坏的。"我们业已看到能够有合理的非理性。同样，能够有道德的不道德或者无过之错。在这样的情形中，不道德的是行动而不是行动者。

它可能再次遭到诘难说："克莱尔造成的坏结果本来是可以避免的。这不像外科医生给予病人最好的可能的治疗时无法避免造成疼痛一样。那个坏的结果是一个独立的和自愿的行动的结果。既然它本来是可以避免的，那么就不能主张说它是最好的可能的后果组中的一个后果组的部分。"

克莱尔会回应说："我本来能够不这样做。但这只是意味着，如果我的动机有所不同的话，我本会这么做。鉴于我的实际动机，我采取不同的行动因果上是不可能的。况且，如果我的动机有所不同的话，这会使结果在总体上说来更糟。既然我的实际动机以后果论者的标准衡量是最好的可能组中的一个，那些坏的后果在相关的意义上是最好的可能后果组中的一个后果组中的部分。"

它可能遭到诘难："如果你采取不同的行动在因果上是不可能的，鉴于你的实际动机，你不能够做出你应该去做什么的主张。应该蕴含能够。"

凯特在第6节中回答过这个诘难。不能够主张说，如果克莱尔本来不能够这么做，她本该采取不同的行动。这个条件从句并不意味着"如果鉴于她的实际动机这本来在因果上是不可能的"，而是意味着"如果无论她的动机会是什么这本来在因果上是不可能的"。

像凯特一样，克莱尔假定心理决定论这一点可能是错的。如果是这样，她的主张可以加以修正。她应当不再主张，如果她具有最好的可能动机组中的一组，这将不可避免地导致她去做她认为是错的事。取而代之的是，她可以主张："如果我是一个纯粹的行善者的话，不去做我认为错的事会是容易的。既然我具有另一组动机，不以这种方式行动就是非常难的。而且，我改变我的动机以便更容易地不以这种方式行动会是错的。既然如此，当我以这个方式行动的时候，我只在非常弱的意义上是道德上坏的。"

接着考虑

事例二。克莱尔要么会救她孩子的命，要么会救数个陌生人的命。因为她爱她的孩子，她救了孩子，而那些陌生人都死了。

在这个事例中，克莱尔会有同样的主张吗？数个人的死亡是一个非常坏的结果。她会主张这些人的死亡是那些最好的可能后果组中的一部分吗？答案可能是否定的。如果克莱尔并不爱她的孩子的话，本来可能会使结果更好。而这对她会更糟，对她的孩子而言则更加糟得多。但是，那么她本来可以救几个陌生人的性命。这个好结果可能会超过那些坏结果，使结果总体上说来更好。

如果是这样，克莱尔会说："我没有什么理由相信，我对我的孩子的爱会有如此坏的结果。我任由我自己爱我的孩子在主观上是对的。而造成自己失去这样的爱是会受责难的，或者从主观上说是错的。当我救我的孩子的命而不是那些陌生人时，我是在按照这样一组动机行动，如果我使自己失去这组动机的话，会是错的。这足以证明我的这样一个主张的正当性，就是当我以这种方式行动时，这是一个无过之错的情形。"

后果论者或许会说："当克莱尔得悉她能救那些陌生人时，她造成自己不去爱她的孩子不会是主观上的错。这将是对的，因为这么一来她会救了那些陌生人。"但是克莱尔会回答说："我本来就不可能像所要求的那样快地失去这种爱。固然存在一些我们能够改变我们动机的方式，但是，就我们那些最深沉的动机来说，这需要一段长长的时间。我力图失去我对自己孩子的爱本来就是错的。如果我力求失去的话，也只有在那些陌生人死去之后才会取得成功。而这只会使结果变得更糟。"

正如这个回答所表明的，克莱尔的主张本质上诉诸某些事实性的假定。如果她具有一个纯粹的行善者的意向的话，本来总体上说会使结果更好，这一点或许为真。但我们是在假定此为假。我们是在假定，如果克莱尔具有某组动机，这组动机有时将使她选择做她认为将

使结果更糟的事，那么结果将更好。而且我们在假定，她的实际动机组是最好的可能动机组中的一组。

我们可以设想其他本来会使结果甚至更好的动机。但是，鉴于有关人性的事实，这样一些动机并非是因果上可能的。既然克莱尔爱她的孩子，那么她会救孩子而不是几个陌生人。我们可以设想每当他人的生命处于危险之中的时候我们对我们子女的爱会“关停”。如果我们都具有这类爱的话，这会使结果更好，这或许为真。如果我们都给拯救更多人的生命以这样的优先权，我们对我们子女的爱会不得不“关停”的情形就会为数不多。因而这种爱会像现在一样多。但是，我们的爱会像这个样子事实上是不可能的。我们不会进行这样的“微调”。当我们孩子的生命遭受威胁的时候，我们对子女的爱不会仅仅因为数个陌生人也受到威胁而关停。[13]

克莱尔主张说，当她做她认为将使结果更糟的事情的时候，她是在错误地行动。但是她还主张：“因为我是在按照这样一些动机行动，就我来说如果失去了这些会是错的，那么这些行动是无过的。当我以这种方式行动的时候，我无需把自己当作坏的。如果心理决定论并不为真的话，我就只在一个非常弱的意义上会是坏的。当我以这种方式行动的时候，我不应当感到懊悔。我们也不应当打算力求再也不以这种方式行动。”

现在可能有人会诘难，既然她作出这样的主张，克莱尔不可能真正相信她是在错误地行动。但是有足够的根据认为她的确具有这个信念。考虑一下她救她的孩子而不是数个陌生人的那个事例。尽管她爱她的孩子，克莱尔并不认为她孩子死亡会是一个比几个陌生人的死亡更糟的结果。她孩子的死亡对孩子和她会更糟。但是数个陌生人的死亡总体上说来会更糟糕得多。在救她的孩子而不是数个陌生人方面，克莱尔是在做她认为将使结果更加糟糕得多的事情。她因而认为，她是在错误地行动。她的道德理论直接蕴含这个信念。她还认为她不应当感到懊悔。但是她据以相信这一点的理由并不在她有关她是在错误

地行动的信念上投下怀疑的阴影。她的理由是，她是在按照一个动机——爱她的孩子——在行动，就她而言造成自己失去这个动机本来就是错误的。这可能表明她并不该当受责难，但是并不表明她不能够相信她的行动是错误的。

这或许可以这样说

> （G4） 如果某个人按照他应该使自己具有的一个动机而行动，而且就他来说使自己失去这个动机会是错的，那么他不会是在错误地行动。

如果（G4）得到正当性证明的话，它会支持有关克莱尔的行动本来就没有错的主张。而且这会支持有关她不能够真正地相信她的行动会错的主张。但在第16节中，我将举一个（G4）并不令人信服的事例。

克莱尔会补充说，在许多别的可能事例中，如果她认为她的行动是错的，她就会认为她自己是坏的，而且她会感到懊悔。如果她去做她认为会使结果更糟的事的话，而且她并不是按照这样一组动机在行动，这组动机因她而失去是错的话，那么情况就常常会如此。后果论一般并不打破一个行动是错的这一信念与谴责和懊悔之间的链接。我们一直在讨论的就是下面这些事例的情形之一：其中某个人按照一个自己失之会错的动机采取行动的那些事例。

还有另一种事例，其中的那个链接是断裂的。C适用于一切，包括责难和懊悔。根据C，当这会使结果更好时，我们应该责难他人和感到懊悔。当责难或懊悔会导致我们以使结果更好的方式改变我们的动机的时候，情形就会如此。当我们像克莱尔那样具有最好的可能动机组中的一组的时候，就不会如此。甚至当我们并不具有这样的一些动机的时候或许也并不会如此。如果我们过于经常地受到责难，责难的后果就会打折扣。C于是可能蕴含着，即便我们并不具有最好的动机组中的一组，我们也应当只为那些我们认为将会使结果变得糟糕得多的行动受到责难。

15.避免错误地行动是不可能的吗?

克莱尔的主张蕴含着，她无法避免做她认为是错的事情。她或许说："我既具有最好的可能动机组中的一组，又决不做我认为是错的事情，在因果上这并不是可能的。但如果我是一个纯粹的行善者的话，我的日常行动就决不会错。然而，在允许自己保持做一个纯粹的行善者的时候，我就会是在错误地行动。如果取而代之的是我造成自己具有最好的可能动机组中的一组，作为我应该去做的，那么我有时会做我认为是错的事情。如果我并不具有纯粹的行善者的意向，那么，我总是像一个纯粹的行善者那样行动，决不做我认为是错的事情，就不是因果上可能的。既然这不是因果上可能的，而且就我造成自己做一个纯粹的行善者来说会是错误的话，那么我不能够因为有时未能像一个纯粹的行善者那样行动就受到道德批评。"

现在可能有这样的说法，正如克莱尔所表述的，C缺乏任何一个道德理论的那些本质特征中的一个。可能遭到这样的诘难："没有什么理论能够要求不可能的事。既然我们不能够避免做C所称为错的事，那么我们就不能够总是做C所主张的我们应该做的事情。因此我们应当拒斥C。如前所述，应该蕴含着能够。"

即便我们否定心理决定论这个诘难也适用。假设克莱尔救的是她的孩子而不是数个陌生人。她本来就会以这种方式行动，因为她并不具有一个纯粹的行善者的意向。她对她的孩子的爱本来就比她的避免做她认为是错误的事情的愿望强烈一些。如果我们否定决定论，那么，我们将否定：在这个情况下克莱尔要避免做她认为是错的事情本来在因果上是不可能的。通过意志的努力，她本来能够违背她的最强烈愿望去行动。即便我们如此主张，我们也不能够认为克莱尔在不具有一个纯粹的行善者的意向的情况下，就总是会像一个纯粹的行善者那样行动。甚至那些否定决定论的人们也不能够完全打断我们的行动和我们的意向之间的链接。

在没有一个纯粹的行善者的意向的情况下，如果我们不能够总是像一个纯粹的行善者那样行动，那么上述诘难就仍然适用。即便我们否定决定论，我们也必须承认存在下面的情况。我们是在假定我们由衷地相信，如果我们都是纯粹的行善者的话结果会更糟。如果我们具有这一信念，我们决不做我们认为将使结果更糟的事情这一点就是不可能的。如果我们使自己或者任由自己保持做一个纯粹的行善者，我们就从而是在做我们认为将使结果更糟的事情。如果我们还有别的愿望和意向，那么我们总是像一个纯粹的行善者那样行动、决不做我们认为将使结果更糟的事情这一点就是不可能的。诘难者因而能够说："即使决定论不为真，我们决不做我们认为将使结果更糟的事情这一点也是不可能的。在主张我们决不应该以这种方式行动方面，C是在要求不可能的事情。既然应该蕴含能够，C的主张就是不可辩护的。"

克莱尔会回答说："在绝大多数情况下，当某个人错误地行动的时候，他该当受责难，而且应当感到懊悔。这就是'应该蕴含能够'学说中最令人信服之处。难以相信，会有这样一些事例，在这样的情形中一个人无论做什么，或者无论早先本会做什么，他都该当受到责难，他都应当感到懊悔。难以相信，某一个人避免以一种遭到责难的方式行动本来就是不可能的。C并不蕴含这个信念。如果我救的是我的孩子而不是数个陌生人，我会认为我是在做将使结果更糟的事情。我会因而认为我是在错误地行动。但这本是一个无过之错的事例。根据C，我们能够总是避免做该当责难的事情。这足以满足'应该蕴含能够'的学说。"

我们可能认为这些主张并未充分地答复这个诘难。对S也有过一个类似的诘难。我们决不做S宣称是非理性的事情这一点是不可能的。我开始答复那个诘难借助的是诉诸第5节中的那个事例：谢林对武装抢劫的回应。在这个事例中，按照任何一个令人信服的有关合理性的理论，我都不免会非理性地行动。要答复对C的这个诘难，克莱尔或许可以诉诸其他一些我们不免会错误地行动的事例。就连最反对

C的一些作者也已经主张说，存在这样的一些事例。在注释[14]中我讨论这个答复。

16.使自己错误地行动可能是对的吗？

既然C是间接自败的，那么它告诉我们使我们自己去做或者更有可能去做它宣称的道德上错的事情。以C的标准衡量这不是一个缺陷。我们能够问一个像我们关于自利论所问的同样的问题。C给我们设定了一个实质性的道德目标：历史尽可能顺利地行进。它还给我们设定了第二个实质性的目标——我们决不错误地行动吗？按照C的那个众所周知的形式——功利主义，答案是否定的。对功利主义者来说，避免做错是达到那个实质性的道德目标的一个纯粹手段。它本身并不是一个实质性的目标。按照C的那些不以一个而以数个道德原则来判断结果的好的版本，也会如此主张。例如，这或许为那个既诉诸功利主义主张又诉诸平等原则的理论所主张。所有这些理论都给我们设定有道德地行动和避免做错的形式上的目标。但是这些理论都会主张这个形式上的目标不是我们的实质性的道德目标的组成部分。

尽管这个主张可能是由任何一个后果论者得出的，但是它不会是按照若干其他道德理论得出的。按照这些理论，避免做错本身就是一个实质性的道德目标。如果我们接受这些理论中的一个，我们可能至少以两种方式诘难C。我们可能说："一个可以接受的理论不能够把道德地行动当作一个纯粹的手段。"这个诘难我在第19节中予以讨论。我们还可能说："一个可以接受的理论不能够要求我们造成自己去做这个理论本身宣称是错误的事情。"

我们应当问，如果我们提出这个诘难，我们自己是否认为所讨论的问题中的行动会是错误的。我们是在考虑这样一些事例，在这些事例中后果论者认为，尽管他是在错误地行动，但他并非是道德上坏的，因为他是按照就其而言使自己失之则错的那些动机在行动。在这些事例中，我们自己认为这个后果论者是在错误地行动吗？

在那个想象出来的克莱尔救她的孩子而不救陌生人的事例中这是不大可能的。如果我们不是后果论者，我们将很有可能认为，克莱尔的行动本来就不会是错的。对此类的其他一些事例我们可能同样这样认为。假设克莱尔忍住不杀我，虽然她真切地相信，杀我会使结果更好一些。克莱尔会想，在忍住不杀我方面，她会是在错误地行动。但是她可以把这个当作一个无过之错的事例来对待。她错误地行动，是因为她强烈地倾向于不杀，而且出于第10节末尾所给出的那个理由，她相信这个意向是一个她若使自己失之则错的意向。我们可能再一次相信，在忍住不杀我方面，克莱尔并不是在错误地行动。

如果这就是我们对有关这些事例所认为的，那么我们应当诘难C的这个部分这一点就愈加不那么明确。我们接受C的这样一个主张，在这些事例中，克莱尔不会表明她自己在道德上是败坏的，或者是该当受责难的。就此没有什么异议。我们可能诘难C的这样一个主张，就是，克莱尔尽管不应该受责难，她的行动却是错的。但是，如果它并不具有它的这些通常的含义，我们恐怕就不应当诘难这个主张。

我们仍然可能诘难说，一个可以接受的道德理论不会告诉我们使我们自己去做这个理论宣称是错的事情。但是请考虑

我的道德堕落。假设我担任某种公职，如果我卷入丑闻的话就会被毁掉。我有一个仇家，他是我所揭穿过的一个罪犯。这个仇家现在获释了，图谋报复。他不是简单地伤害我，而是决定迫使我自甘堕落，因为他知道我将把这看得比最重的伤害还要更严重。他威胁我说，除非我同意他给我拍摄淫秽照片，否则他本人或者他帮会中的人会杀死我所有的子女。如果他后来把这些照片寄给某个新闻记者的话，我的公职生涯就会毁掉。他然后就能够通过威胁要毁掉我的公职生涯而迫使我选择错误的行动。他将迫使我选择帮助他犯罪。尽管我在道德上与绝大多数人一样好，但我不是一个圣人。单单为了保住我的职业生涯我不会非常错误地行动，但是我会帮助我的仇家犯一些小罪。即便有鉴于我倘若拒绝帮助我的仇家我的职业生涯就会被毁掉这个事实，我也会错误地行动。我们下面能够假设，既然我非常了解我的仇家，我有充分的理由，既相信倘若我拒绝让他拍照，我的子女会被杀掉；又相信如果我不拒绝，我的子女将不会被杀掉。

我应该让这个人拍照。我们似乎能够令人信服地主张，政府不应当屈服于这样的威胁，因为这只会使政府面临随后进一步的威胁。但是这个主张并不包括我的仇家对我的这个威胁。就我而言，在预见到我的子女结果将被杀掉的情况下，拒绝他的要求会是错的。我应该让他拍照，即便我知道后果将是我随后将时常犯错误。在我的子女获得安全之后，我将时常为了保住我的职业生涯而帮助我的仇家犯一些小罪。后来的这些行动将是相当自愿的。我不能够主张说，是我的仇家后来的那些威胁迫使我以这些方式行动。我本可以拒绝错误地行动，即便这会毁掉我的生涯。

我业已主张，我应该让这个人拍照。这甚至会得到绝大多数拒斥后果论的人们的认同。这些人会同意，既然那是救我的子女的惟一方式，我应该使我以后将时常错误地行动这一点属实。这些人于是认为，一个可以接受的道德理论应能够告诉某个人要让自己做这个理论宣称是错误的事情。既然他们相信这一点，他们就不能够诘难后果论能够有这个蕴含。

如果我让我的仇家拍照，我会变得倾向于帮助他犯一些小罪。且让我们给这个事例增添某些特征。通过放弃我的职业生涯，我本来能够使自己丧失这个意向。但是我的仇家已经威胁说，如果我不按他意志去做，他的帮会将杀掉我的子女。因此，我使自己丧失这个意向会是错的。相反，如果我拒绝帮助我的仇家犯罪，他通过把我丑态百出的淫秽照片寄给记者，毁掉的将只是我的职业生涯。我的仇家向我保证，如果他毁掉我的职业生涯，他将不会杀死我的子女。他通过威胁毁掉我的职业生涯，从造成我做我相信是错的事情中得到邪恶的快感。如果他的威胁是杀死我的子女的话，这种快感就会失去。如果我帮助他犯罪是因为这是保住我的子女生命的惟一途径，我就不会认为我是在错误地行动。因为我的仇家想要我认为我是在错误地行动，所以他并未实施这个威胁。

由于了解我的仇家，我有充分理由相信他所说的话。既然是救我子女生命的惟一方式，我应当让他拍照。我应该使我自己倾向于帮助他去犯小罪。而且，就我而言使自己丧失这个意向会是错的，因为如果我丧失了的话，我的子女将遭到杀害。但是，当我按照这个意向去行动的时候，我是在错误地行动。我不应该帮助这个人去犯罪，只是出于保住我的职业生涯。

这个事例表明，我们应当拒斥我所称的（G4）。这是这样一个主张，如果我应该使自己具有某个意向，而且就我而言使自己丧失这个意向会是错的，那么当我按照这个意向行动的时候就不会是在错误地行动。在刚刚表述的那个事例中，当我按照这样的一个意向行动的时候，我是在错误地行动。[15]

现在我将一并陈述四个类似的错误。一些人主张，如果我使自己具有某个意向是合理的，那么按照这个意向行动就不能够是非理性的。这一主张通过那个我所称的谢林对武装抢劫的回应事例被表明是错谬的。第二个主张是说，如果就我使自己相信某个行动是合理的这一点而言是合理的，那么这个行动就是合理的。这个主张被我所称的“我的奴隶状态”的事例表明是错谬的。第三个主张是，如果存在某个我应该使自己具有的意向，而且如果我使自己失去这个意向会是错的，那么就我来说按照这个意向行动就不能够是错的。刚刚举出的事例表明这一主张是错谬的。第四个主张是说，如果我应该使自己相信某个行动不会是错的，那么这个行动就不能够是错的。我将很快予以表明该主张是错谬的。这四个主张假定，合理性与正确性能够被遗传或被转移。如果我使自己倾向于以某个方式行动或使自己相信这个行动是合理的或正确的这一点是合理的或正确的，那么这个行动就是合理的或正确的。我所举出的事例表明这并非如此。合理性与正确性是不能以这种方式遗传的。在这个方面真理比这些主张所蕴含的要更为简洁。

17.何以后果论可能是自谦的

或许有人主张说，如果后果论有时打断我们的行动是错的信念与我们自己是坏的信念之间的链接，那么我们事实上就不会继续足够认真地对待道德。同样，如果我们相信其他愿望应当经常更强一些的话，那么我们避免做错的愿望就可能遭到逐渐削弱。只有当我们相信它应当总是压倒性的，而且当它倘若不是压倒性的话我们则感到懊悔的时候，这个愿望才可能幸存下来。或许有主张说，按照这些或其他根据，如果我们总是保持我们的道德信念与我们的意图和情绪之间的链接，会使结果更好。如果是这样，如果我们并不相信C的话，那么这会使结果更好。

我怀疑这些主张。但是值得考虑一下它们会蕴含什么。根据C，我们每一个人都应当力求具有以后果论者的标准衡量的最好的可能愿望和意向组中的一组。如果我们并不仅仅具有这些愿望和意向，而且还具有相应的道德情绪和信念的话，那么这可能使结果更好一些。

考虑一下，比如，盗窃。按照C的某些版本，如果财产被盗窃，那么这无疑是坏的。按照C的另一些版本，则并非如此。按照这些版本，只有当盗窃使结果更糟时，它才是坏的。避免盗窃并非是我们的终极道德目标的组成部分。但是，如果我们强烈地不倾向于偷盗的话，会使结果更好，这可能为真。而且，如果我们相信偷盗无疑是错的，在偷盗时会感到懊悔，那么这或许会使结果更好。类似的主张可能适用于许多其他种类的行动。

如果这些主张为真，C会是自谦的。C会告诉我们说，我们应当力求相信的不是它自己而是某个别的理论。我们应当力求相信这样一个理论，如果我们相信它的话，结果会最好。按照上述那些主张，这个理论或许不是C。它可能是西季威克所称的常识道德的一个版本。

如果C告诉我们相信这个道德的某个版本的话，这个版本就不会是像现在这个样子的常识道德，而是一个改进版本。常识道德不是那个会使结果最好的道德理论信念。例如，这样的一个理论会对富人要

求得更多。如果那些富裕国家的人们送给穷人至少年收入的四分之一或一半，或许使结果最好。富人现在只把不到百分之一的收入送给穷人，而且看起来相信他们这样做是正当的。

假设C告诉我们相信某个别的理论。正如我已经说过的，如果据以改变我们信念的理由不是一个给我们的旧有信念投下怀疑阴影的理由，而只是倘若我们具有不同的信念就会有好的后果的话，那么，在这样一种情况下改变我们的信念会是一件困难的事。但是我们有多种多样的方式可以带来这种变化。或许我们都会对此着迷，而下一代人则以不同方式加以培养起来。我们得设法忘掉我们是如何和为何获得我们的新信念的，而且该过程得设法向未来的历史学家们隐藏起来。

如果我们接受的不是C，而是集体性的后果论，那么，这里会有差异。如果我们接受C，那么，我们或许得出结论说，C应该为绝大多数人所拒斥，但是应当仍然为一些人所相信。我们的理论从而会是部分自谦的，而且是部分隐秘的，因为它告诉那些相信它的人不要启蒙无知的大多数人。但是作为集体性的后果论者，我们应当相信这样一种道德理论，就是，如果我们都相信它的话，结果会最好。这个理论不能是隐秘的。

一些人发现，一个道德理论或许是隐秘的这一点尤其值得诘难。如果我们相信欺骗在道德上是错的，那么有关道德的欺骗可能看起来尤其是错的。西季威克写道："如果本身保持隐秘的话，那么隐秘性的道德是权宜性的这一学说看起来就是权宜性的。或者，如果这种隐匿难以维持的话，那么这样一点可能是可取的，就是常识应当批判这样一些学说，它们局限于启蒙过的少数人是权宜的。"[16]这就是威廉斯所称的"政府大厦"后果论，因为它把大多数人当作殖民地中的土著那样看待。[17]正如威廉斯所主张的，我们不能够欢迎这样的一个结论。西季威克对他的结论感到遗憾，但是他并未把遗憾看作是怀疑的一个根据。[18]

我已经主张，C完全自谦是不大可能的。它充其量是部分自谦，部分隐秘的。如果某些人并不相信C的话，或许会使结果更好一些；但是，如果C没有一个人相信的话，会使结果更好这一点就不大可能了。

再举一个怀疑这一点的根据。假设我们都最终相信C。（当我们记起C可以是一个多元理论，诉诸许多不同的道德原则的时候，这看起来将较少令人怀疑。）然后我们断定C是完全自谦的。我们断定，如果我们使自己相信常识道德的某个改进版本的话，会使结果最好。我们可能成功地促成我们信念中的这个变化。鉴于世界的变化和技术的变革，如果我们修正我们的道德信念的话，结果会更好，这一点或许后来会实现。但是如果我们不再相信C，因为我们现在相信的是常识道德的某个版本的话，我们就不会被引导在我们的道德中去进行所需的这些修正。我们据以相信这个道德的理由不会是：我们现在相信它本身是使结果最好的那个道德信念。这就是为什么我们促使我们自己相信这个道德的理由。但是为了相信这个道德，我们必须忘记这就是我们所做的。我们现在只会相信这个道德。因而我们可能不会被引导去修正我们的道德，即便我们对这个道德的信念真会增加某种灾难的机会，诸如核战争。

这些主张应当影响到我们对于倘若我们都不再相信C的话是否会使结果更好这一问题的回答。我们可能会正确地认为存在某个别的道德理论信念，从短期来看会使结果更好。但是一旦后果论自谦，绳索被切断，长期的后果可能会糟得多。

这表明，最可能会为真的是，C是部分自谦的。如果借助某种想成功必须予以忘记的自欺过程，绝大多数人使自己相信某个别的理论，可能更好一些。但是作为一种预防措施，一些人应当继续相信C，而且应当保留关于这种自欺的有说服力的证据。这些人无需生活在“政府大厦”中，或者具有任何别的特殊地位。如果一切进展顺利的话，这些为数不多的人不会做什么。但是，如果绝大多数人所相信的

那个道德理论的确变成灾难性的话，那些为数不多的人就会出示他们的证据。当绝大多数人得悉他们的道德信念是自欺的结果的话，这就会从根本上削弱这些信念，并且防止这种灾难。

尽管我已经说过，假设C是完全自谦的话，这未必可能。假设C告诉我们大家都使自己不相信它自己，而是相信某个别的理论。威廉斯主张，如果是这样，该理论就不配享用这个名称，因为它"一点都决定不了思想在世界中如何被实施"[19]。这个主张令人不解，因为，就像威廉斯还主张的那样，C对我们思考道德的方式会有所要求，而且我们的愿望或意向组"必须是追求最好的"。[20]这是在要求某个相当具体的事情，而且是完全后果论的。

威廉斯提出第三个主张，如果C是完全自谦的话，它会不再有效。[21]这无需如此。假设事情如上述的那样发生。我们都最终相信C的某个形式。然后我们由衷地相信，如果我们都相信某个别的理论的话，这会产生最好的可能结果。C则告诉我们都要相信这个别的理论。以某种间接的方式，我们促成我们信念中的这个变化。现在没有什么人相信C。但这并不证明C已经不再生效的主张是正当的。它有这样一个后果，就是我们现在都相信某个别的特定理论。而且我们对这个别的理论的信念将产生最好的可能结果。尽管没有什么人相信C，C仍然是有效的。有两个持续的事实就是我们早先相信C而产生的后果：一个事实是我们的那些新的道德信念，另一个事实是，因为我们具有这些信念，结果就会尽可能地好。

威廉斯正确地主张，如果C是完全自谦的，它表明什么就会不清楚。我们会不得不决定它表明的是C"不可接受呢，还是仅仅表明没有什么人应该接受它"[22]。清楚的是，按照我们最后的那些假定，没有什么人道德上应该接受C。如果有什么人的确接受C的话，它自己会告诉他，他道德上应该拒斥C，而且代之以相信某个别的理论。但是，正如威廉斯所提出的，存在着两个问题。一个问题是，某个理论是否是那个我们道德上应该力求相信的理论。另一个问题是，该理论是否

是那个我们理智上应该或者以追求真理的标准衡量应该相信的理论——该理论是否是真的或者得到最好证明的理论。我早就主张，如果一个关于合理性的理论是自谦的，这并不表明这个理论不能是真的或不能是得到最好证明的理论。我们能够就道德理论做出类似的主张吗？

我们对这个问题的回答部分取决于我们有关道德推理的本质的信念。如果一个道德理论能够是完全直截了当地真，那么显然，如果它是自谦的，这并不表明它不能够是真的。但是我们可能代之把道德当作一个社会产物，要么是以事实上的方式产生的，要么是以某种"理想的建构主义者"方式产生的。我们可能然后主张，要成为可以接受的，一个道德理论必须满足罗尔斯所称的"公开性条件"：它必须是一个人人应该接受的理论，人人彼此公开承认的理论。[23]按照这些元伦理学的观点，一个道德理论不可能是自谦的。按照其他一些观点，它却是能够自谦的。在这些不同的观点之间作出决断至少要占用一本书的篇幅。因此，在本书中对这个问题我们只能存而不论。如果像我所认为的，C不会是自谦的话，这无关紧要。

18.对假定不变性的诘难

我现在回到早先提出的一个诘难。考虑一下那些主张自利论间接自败的人。假设这些人相信S，而且结果从不克己。对他们而言这样更糟。如果他们具有某些别的愿望或意向的话对他们会更好。就这些人的情况而言，除非他们相信一个不同的理论，否则这不会是可能的。而且，他们既不能够改变他们的信念又不能改变他们的意向，这可能是真的。

就后果论者而言类似的主张或许为真。假设，因为我们都相信C，我们都是纯粹的行善者。这使结果比如果我们都具有别的意向时要更糟。我们才能够改变我们的意向，除非我们改变我们有关道德的信念，但我们不能带来这些改变。

所有这些主张未必为真。如果它们为真的话，它们会提供对S和C的诘难吗?

考虑一个想象的事例可能有所帮助。假设撒旦统治宇宙。撒旦不能够影响到哪一个是有关合理性的真理论，或者哪一个是最好的或得到最佳证明的理论。但是他知道这个理论是哪一个，而且他乖戾地造成凡信仰这个理论就会具有以这个理论自身的标准衡量的坏的后果。在设想这个事例方面，我们无需假定最好的理论是那个自利论。无论最好的理论是什么，撒旦都会造成凡信仰这个理论就会具有以这个理论的标准衡量的坏的结果。

其次就道德理论我们也能够作同样的假定。假设最好的道德理论是功利主义。照此理论，我们都应当力求产生那个不偏不倚地考虑的、对每一个人都最好的结果。撒旦使人们确信，如果相信这个理论，那么对每一个人来说则会更糟。再假设最好的道德理论不是后果论的，而且它告诉每一个人绝不要欺骗他人、强迫他人或不公正地对待他人。尽管与他们的意向相反，但是撒旦使那些相信这个理论的人们事实上更加具有欺骗性、强迫性和不公性。

撒旦使人们确信，如果任何一个人相信某个理论，那么这会带来以这个理论的标准衡量的糟糕的后果。这会表明这样一个理论不是最好的理论吗？显然不会。它充其量表明的不过如此：鉴于撒旦的介入，如果我们并不相信那个最好的理论的话，它会更好。既然我们纯粹是撒旦玩弄于股掌之间的玩物，有关实际的真相就是极端令人压抑的。如果我们也不知道这个真相或许会更好。

在这个想象出的事例中，如果我们并不相信这种最好的理论的话会更好一些。这表明我们应当拒斥

（G5） 如果我们应该造成自己相信某个行动是错的，那么这个行动就是错的。

正如我所主张的，错误性并不以这种方式遗传。

假设我们所知道的一切，只是相信某个理论会具有以这个理论的标准衡量的坏后果。这并不能表明这不是最好的理论。是否如此，必定取决于为什么相信这个理论则具有这些糟糕的后果。有两种可能性。这些糟糕的后果可能产生自我们成功地做了这个理论所告诉我们要做的。如果这是真的，那么该理论就会是直接地自败的，而且这可能就驳倒了这个理论。这些糟糕的后果可能并不这样产生，而是由某个关于实在的完全分离的事实所造成。如果这个事实是由于撒旦的介入，这就不会给该理论投下什么怀疑的阴影。

就上述可能性我们应当主张什么？假设下述情况为真。对我们每一个人来说，如果相信S并因而是永不克己的，结果就会更糟。如果我们都相信C，并因而是纯粹的行善者，那么这会使结果更糟。而且，如果我们具有这两个信念和意向中的一个，我们就不能够改变它们。那么相信这两个理论会具有以这些理论的标准衡量的坏的后果则会是真的。这会给这些理论蒙上怀疑的阴影吗？抑或它只不过是像撒旦那样的干预？

最好的理论可能既不是S也不是C。我将在更后面论证我们应该拒斥S。但是，如果我是错的，而且要么S要么C是最好的理论，那么，我提出，上述的可能性就不会提供对其中任何一个理论的诘难。如果要么S要么C是最好的理论，那么，相信这个理论就会具有以这个理论的标准衡量的坏的后果。但是这些坏的后果并不会是我们去做或者力图去做S或C所告诉我们去做的那些事情的结果。那些坏的后果会是我们的意向的结果。而且这些理论不会告诉我们要具有这些意向。它们会告诉我们，如果我们能够，就不要这么做。S会告诉我们，如果我们能够，就不要永不克己。C会告诉我们，如果我们能够，不要做纯粹的行善者。我们之所以会有这些意向中的一个意向，是因为我们信仰这些理论中的一个理论。但是这些理论并不告诉我们要去相信它们自己。S告诉每一个人要相信那个会对他最好的理论信念。C告诉我们

要相信那个会使结果最好的理论信念。按照上面所作的假定，S和C会告诉我们不要相信S和C。

因为我们要么相信S要么相信C，信仰其中任何一个理论都会具有以此理论的标准衡量的坏的后果。但是这些坏的后果并不会是我们做这些理论告诉我们所做的事情的结果。它们会是我们具有一些意向的结果，这些理论并没有告诉我们要去具有这些意向，而且告诉我们倘若能够的话不要具有。而且它们会是这样一种结果，那就是，我们所相信的并不是这些理论告诉我们所相信的，而是它们告诉我们倘若能够的话所不要相信的。既然这些坏后果不能够以这些方式中的任何一个方式归咎于这些理论，所以我提出，如果上述所提出的主张为真的话，这不会在这些理论上面投下怀疑的阴影。像撒旦的干预那样的，这些主张只会是关于实在的一些令人压抑的真相。

19.合乎理性或合乎道德会是一种纯粹的手段吗？

S告诉我们要合乎理性地行动，而C告诉我们要合乎道德地行动。但是这些只是我所称的我们的形式上的目标。我已经假定，合乎道德地行动本身不会是C给我们设定的一个实质性的目标。C或许主张，合乎道德地行动是一种纯粹的手段。同样，合乎理性地行动可能不是S给我们设定的那个实质性的目标的组成部分。而且S可能主张，合乎理性地行动是一个纯粹的手段。这是对这两个理论的诘难吗？

此处在S和C之间存在一个差异。S不能够主张，我们的形式上的目标本身是一个实质性的目标。但是C本来能够提出这个主张。在此对S可能存在一个诘难。但是对C却并不能存在一个类似诘难。

我们或许对S诘难说："如果合乎理性地行动不是我们应当具有的一个目标，而是一种纯粹的手段，那么我们为什么应当合理地行动？我们为什么应当知道什么是我们最有理由做的？"

一个自利论者应当如何回应？他或许采纳有关自利的目标列表理论。那么他或许主张："合乎理性和合乎理性地行动它们本身是那使我

们的生活变得更好的东西的组成部分。如果它们总体说来对我们更好，那么S并不蕴含着合乎理性和合理行动是纯粹的手段。它们本身就是每一个人终极的T设目标的组成部分。”

下面考虑一个本身是快乐主义者的自利论者。这个人必须承认，他相信合乎理性地行动是一种纯粹的手段。但是他也许会说：“根据S，你最有理由做的事情是凡是可以使你的生活过得尽可能好的事情。如果你想知道你最有理由做的是什么，而且想要合乎理性地行动，S并不蕴含着这些是毫无意义的愿望。这并不为这样一个主张所蕴含，就是，如果你遵循S，而且合乎理性地行动，你的行动的重要程度只是一种手段。作为一种手段的重要程度也是一种重要的方式。只有当合理行动并不重要的时候，你的愿望才会是毫无意义的。S主张，当你在决定做什么的时候，与合理行动这一点相比，没有什么更为重要。即便对你而言做来合理的是使你自己在意向上倾向于非理性地行动，这最后一个主张也是得到正当证明的。即便此处，最为重要的也是做那对你而言做来会是合理的事。”

我现在从S转到C。C或许主张合乎道德地行动是一个纯粹的手段。我们可能诘难说：“如果是这样，我们为什么应当在乎道德？”首先考虑后果论者中最简单的一种——快乐主义的功利主义者。这样的一个人或许说：“所发生的是好还是坏，在道德上是重要的。如果受苦更多则是坏的，如果幸福更多则是好的。我们的行动是否是道德的，我们是否避免做错，也是重要的。我们应当尽自己最大的努力减少苦难，增加幸福。这之所以重要，不是因为它本身，而是因为其后果。在这个意义上，避免做错纯粹是一种手段。但这并不蕴含着，我们是否避免做错在道德上并不重要。当我们在决定做什么的时候，与避免做错相比，没有什么更为重要。即便当我所应该做的是使我们自己倾向于错误地行动，最后这个主张也是得到正当证明的。即便此处，最为重要的也是我们做我们所应该做的。”

一位快乐主义的功利主义者必须承认，按照他的观点，如果做错并不具有坏的后果，那么它就不会是重要的。如果更多地做错，这本身不会使结果更糟。就像S的快乐主义版本的情况一样，我们的形式上的目标的达到的确重要，但只是作为一种手段。

这位功利主义者能够为这个主张辩护吗？他或许首先诉诸威廉斯所谓的道德的自我放纵的无吸引力。[24]比较两个正在力图解除他人苦难的人。第一个人之所以如此行动，是因为他同情这些人。他还相信，苦难是坏的，应该予以解除。第二个人之所以如此行动，是因为他想当一个有道德的人。就这两个人而言，第一位看起来更好一些。但是第一个人对合乎道德地行动的好和做错的坏并没有什么想法。他的行动受到的驱使，只是他的同情，以及他的这样一个信念，就是既然受苦是坏的，他就应该力求避免它。这个人看来把合乎道德地行动当作一个纯粹的手段。正是第二个人把合乎道德地行动当作一个本身就是好的独立的目标。既然第一个人看起来更好一些，这支持有关合乎道德地行动是一个纯粹的手段的主张。

下面考虑一下

> 谋杀和意外死亡。假设我知道X要死了，而且作为他的最后行动他打算谋杀Y。我还知道，除非Z获救，否则他将被森林大火烧死。我或许能够劝说X不要谋杀Y，或许代之能够救Z的命。假设我相信，如果Y被X所杀的话，这不会比Z被森林大火烧死更糟。这两个结果同等的糟糕，因为它们哪一个都包含着某个人丧命。在第一个结果中，还会存在一个非常严重的做错情形。但是，根据我的理论，这并不会使结果更糟。（如果做错者自己生存下来的话或许如此。但是X要死了。）既然我认为如上所述的做错本身并不使一个结果更糟，那么我认为，如果我救Z的几率会略微高过劝说X不要杀Y的几率的话，我应该力求救Z。[25]

许多人会接受最后这个结论。他们会相信，如果我救Z的几率略微高一些的话，我就应当力求救他而不是救Y。如果我们接受这个结论，我们还能够主张，如果有更多的做错的话，它本身是坏的吗？甚至因为它包含着做错，X谋杀Y会是一个更糟的结果，我们仍然能够主张，我们应该力求制止Z的意外死亡吗？为这个主张辩护会是困难

的。对很多人来说，这又是一个支持这样一个观点的事例，就是避免做错纯粹是一种手段。

可能遭到这样的诘难：“如果X打算谋杀Y，坏的本性已经呈现。如果你只是劝说X不要实施他的谋杀意图，你并未防止那个坏的本性。这就是你为什么应该力求救Z的缘由。”我们可以对这个事例做些改变。假设我知道X可能很快错误地意识到他已经被Y所出卖。X并不是道德上坏的，但是他像奥赛罗一样。他是一个好人，却潜在地是一个坏人。我知道，如果X相信他已经被Y所出卖的话，很有可能他将像奥赛罗一样谋杀Y。我阻止X获得这样的错误信念有很高的几率，从而避免他谋杀Y。但是我救下Z的性命的几率更高一些。如前所述，X在任何情况下都是要死的。许多人会再次认为我应该力求救Z的性命。这暗示着，如果X形成他的意图，然后谋杀Y，这不会是一个比Z意外丧生更糟的结果。如果这本来会是一个更糟的结果的话，我为什么应当救Z而不是Y？当我成功的几率只是略微高一些的时候，两恶相权我为什么应当力求制止罪恶较轻的那一个？

假设，因为我们相信我应该力求救Z，我们同意Y之死不会是一个比Z之死更糟的结果。如果是这样，Y之死体现出来的坏的本性就只是Y的死亡。X形成他的谋杀意图然后非常错地行动这一点并不能够使这个结果更糟。正如我所主张的，我们可能得出避免做错纯粹是一种手段的结论。

在这样一些情况下，很多人认为做错并不使结果更糟。如果道德上正确的行动更多一些、得到履行的义务更多一些，这会使结果更好吗？我可能时常应允去做我无论用什么办法都决意要做的事情。从而我会使更多的义务得到履行这一点为真。但是没有人认为这会使结果更好，或者认为这会是我应该做的事情。

接着假设：贫穷被消除，天灾不再发生，人们不再遭受身体上或精神上的疾病之苦，而且在许多别的方面不再需要他人的帮助。这些变化都会在某种程度上是好的。它们会在任何情况下是坏的吗？自己

付出相当的代价来帮助那些落难的人在道德上是值得称道的。在我所描述的这样一个世界中，极少有人需要这种帮助。这些道德上值得称赞的行动就会更少得多。这一点会是坏的吗？这会使这个结果在某种程度上更坏吗？

如果我们的回答是否定的，这再一次支持有关合乎道德地行动是一种纯粹的手段的主张。但是我们中的一些人会给出肯定的回答。我们会认为在这方面结果会更坏。而且还有很多人对错做本身持有不同的观点。这些人会认为，与Z的意外死亡相比，X谋杀Y作为一个结果会坏得多。一个后果论者能够接受这些主张吗？

这取决于他所接受的原则是什么。首先考虑一个快乐主义的功利主义者。如果X谋杀Y不会导致比Z的意外死亡更多的苦难或损失更多的幸福，那么这个功利主义者不会主张，作为一个结果，那个谋杀更坏。当转向道德上值得称赞的行动的时候，下面则是他所能够主张的全部：一个人深信自己在一些重要的方面对人有所帮助是幸福的一个主要源泉。因此，如果只有为数极少的人需要这样的帮助的话，那么这在某种意义上是糟糕的。

其次考虑一个接受有关自利的目标列表理论的后果论者。按照这个理论，合乎道德和道德地行动对我们来说本身可能就是好的，而无论它们的后果可能是什么。它们可能属于那些对我们来说最好的东西，或属于使我们的生活过得最好的东西。而且道德上坏本身对我们来说可能就是最坏的东西之一。如果一个后果论者提出这些主张，那么他就会否定道德地行动和避免做错是纯粹的手段这一点。按照C的任何一种令人信服的说法，如果我们的生活过得好，那么就更好。按照刚刚提出的那些主张，道德地行动和避免做错是C给我们设定的终极道德目标的组成部分。

照此观点，尽管这些是这个目标的组成部分，但是它们本身不是终极目标。它们之所以是这个目标的组成部分，是因为就像幸福一样，合乎道德是使我们的生活过得更好的因素之一。

一个后果论者会有一个不同的主张。他会主张，我们如此这般的形式目标是一个实质性的目标。他会主张，如果有更多的做错，即便这不会对什么人来说更坏，也会是坏得多的。同样，如果更多的人合乎道德地行动，即便这不会对什么人来说更好，也会是好得多的。一个后果论者甚至会主张，达到我们的形式上的目标与达到我们的其他道德目标相比具有绝对优先权。他会接受红衣主教纽曼的观点。纽曼认为，痛和罪两者都是坏的，但是罪绝对坏得多。即使整个人类遭受"最极端的极痛"之苦，其坏的程度也不及犯下可赎的小罪。[26]

极少有后果论者会走得如此之远。但既然纽曼的观点是C的一个版本，我们不能够主张C赋予避免做错的权重太小。与所有其他道德目标相比，C会赋予这个目标以绝对优先权。

再一次看来，C不是一个与众不同的道德理论，但却能涵盖所有的理论。这并非如此。非后果论者可以主张，并非是C赋予避免做错的权重太小，而是C赋予这个权重以错误的方式。按照C的这个极端版本，避免做错是我们共同的道德目标之一。一个非后果论者会说，我们不应该错误地行动，即便我从而使其他人做错的事减少许多。按照C的这个版本，在这种情况下我不会是在错误地行动。如果我是在做最有效地减少做错发生率的事，我就是在做我应该做的事。

20.结论

我现在将对本章后半部分加以总结。我假定，以我所表述的那些方式，后果论是间接地自败的。如果我们始终倾向于做会使结果最好的事情，那么会使结果更糟。如果我们都具有这一意向的话，鉴于人们实际上的情况，结果可能比实际上要好一些。但是，结果会比我们倘若具有某些别的因果上可能的动机组的话更糟。

我问了当C间接自败时它是否以其自身标准衡量是失败的。如果我们总是倾向于做会使结果最好的事，而假使会使结果更糟，那么C

告诉我们不应当具有这个意向。既然C提出这个主张，那么它以其自身的标准衡量不是失败的。

假设我们都接受C。我们的理论告诉我们，我们应当使我们自己具有或保持以后果论的标准衡量的最好的可能动机组中的一组。既然C是间接自败的，那么这一点蕴含如下：如果我们具有最好的可能动机组中的一组，那么我们有时将根据我们自己的理论有意地错误地行动。但是鉴于我们何以错误地行动的那个特殊理由，我们如此行动的时候无需把我们自己当作道德上是坏的。我们能够认为这些是无过之错的一些情况。我们能够相信这一点，是因为我们是在按照就其而言造成自己失之则错的一组动机在行动。

即便C并不是间接自败的，这些主张中的某一些或许也是蕴含其中的。如果C是不切实际地要求的话，这些主张就会蕴含其中。这很可能是真的。很有可能的是，即便我们都相信C，我们变得总是倾向于做我们认为会使结果最好的事情这一点也会是因果上不可能的。如果这是真的，那么C告诉我们力求具有那些最好的可能动机组中的一组。

在提出这些不同的主张方面，C是前后一致的。也并非未能严肃对待道德。即便我们采纳这些主张，仍然会有许多其他事例，我们会把我们自己当作道德上是坏的。每当我们有意使结果糟得多，而不这样做是因为我们具有最好的可能动机组中的一组的时候，情况就会如此。尽管我们不会只在乎避免做错，因为我们还会在乎具有最好的动机，但是我们仍然会把很多行动看作表明了该行动者道德上是坏的。

可能有诘难说，这些主张错误地假定了心理决定论。如果一个后果论者接受这个诘难，他必须限定他的主张。他能够主张，如果我们具有最好的动机组中的一组，对我们来说避免做我们认为是错的事情常常会是艰难的。他必须承认，在这些情况下，我们并非是完全无过的；但我们只是在非常弱的意义上是坏的。

另有一个诘难说，一个可以接受的理论不能够告诉我们要使自己做这个理论宣称是错的事情。但是我举了该诘难甚至会被那些拒斥C的人们所否定的一个事例。

第三个诘难是，既然C是间接自败的，那么我们不能总是避免做C宣称是错的事情。既然我们不能总是做C所主张的我们应该做的事情，C要求的就是不可能的事情。它违背了应该蕴含能够的学说。我认为，对这个诘难可以作出答复。

第四个诘难是，如果我们具有与C相冲突的道德信念，将使结果更好。如果这是真的，C就会是自谦的。它会告诉我们去相信并非它自己，而是去相信某个别的理论。我怀疑这是否为真。我认为，C充其量是部分自谦的，部分隐秘的。如果某些人相信某个别的理论的话，这或许使结果更好，但是如果没有什么人相信C的话，就不会使结果更好。而且，即使C是完全自谦的，我相信这也不会给C投下一层怀疑的阴影。是否如此取决于关于道德的本质和道德推理的最佳观点是什么。既然我还没有论证这些观点中的任何一个，我就并未完全辩护我的这样一个信念：如果C是自谦的话，这不会给C投下一层怀疑的阴影。

最后我问了我们是否能够接受C有关合乎道德地行动纯粹是一种手段的主张。如果我们不能够，C就无需提出这个主张。它甚至会主张，制止做错对于我们的其他道德目标具有绝对的优先权。功利主义者们不能够提出这个主张。但是后果论者们却能够。

在本章的前半部分，讨论了关于合理性的自利论。在本章的后半部分，讨论了那组后果论的道德理论。主张所有这些理论是间接自败的这一点是可信的。而且它们可能还是自谦的。但是，就这些理论的情形而言，间接地自败并不是损害性地自败。这些事实也并不提供对于这些理论的独立诘难。这些事实并不表明，这些理论要么是错误的，要么是得不到辩护的。这可能是真的，但是迄今的论证没有表明

这一点。它们充其量表明，可正当地予以主张的，比我们可能本来希望的，要复杂得多。

---

[1]另有一个例子，请参见Adams（2）的开始部分。

[2]如果对于某人而言存在一个去做X的理由，那么我要说这个人就有一个去做X的理由。按照这些词的这种用法，我们能够有一些我们并未意识到的行动理由。

[3]参见Schelling（1）。

[4]参见Nagel（1），第五章。

[5]Gauthier（3）和（4）之中作此主张。

[6]Sidgwick（1），第386页，脚注（4）。

[7]这些诘难是由S·卡根向我提出的。

[8]我可能要努力为这个论证辩护。我或许说："为什么S告诉我要去相信不顾这些威胁总是合理的呢？只是因为对于我而言具有这样的信念可能要更好一些。在我遇到坏运气，而你发出了威胁之后，这个信念对于我而言便不再是好的。S会告诉我要抛弃这个信念。这个事例因而并不表明我所主张的。对于我而言不顾你的威胁是非理性的。但是当我不顾你的威胁的时候，S不再告诉我去相信这是合理的。我们因而能够保持这样一个主张，就是，如果S告诉我相信一个行动是合理的，那么这个行动就是合理的。我所设想的事例并不提供一个反例。当我们从威胁转向承诺的时候，我们就能够主张我所否定的。既然S告诉我们要相信守诺是合理的，纵使当我们知道对于我们而言这将更糟的时候，信守这样的诺言也是合理的。"
这个辩护是失败的。为什么S告诉各人要具有这个信念？因为，如果他被人们知道不相信这一点的话，他就将被排除在我所描述过的那种协议之外：那些协议尽管是互惠的，但是要求克己。假设我具有这个信念，而且我被接受这样的一些协议。尽管保护我的名声对于我而言时常会更好，但是将存在一些情形，在这些情形中倘若遵守诺言对于我而言则将更糟。当此点为真的时候，S告诉我要丢弃我的那个有关遵守诺言总归合理的信念。我们不能够主张说，对于我而言遵守这些诺言是合理的，因为S告诉我要相信这一点。当我遵守这些诺言时，S不再告诉我要相信这一点。

[9]参阅Adams（2）。

[10]这一点在Sidgwick（1）第431到439页，以及Adams（2）通篇都有论证。亚当斯讨论了他所谓的行动和动机功利主义，主张动机功利主义并不就是行为功利主义的一个特殊情况。一个动机功利主义者会主张，每人都应当拥有那些一旦拥有则将使结果最好的意向，即便这个人在造成自己具有这些意向方面本来就什么也做不到。亚当斯在主张动机功利主义不同于行为功利主义这个方面是对的，就如同规则功利主义至少在某些版本中并非行为功利主义的一个特殊情况一样。对于这些区分，在此我不予讨论。

[11]Mackie第四章，第2节。

[12]集体性的后果论并不是里根所提出的合作后果论，也不是我在第三章中所讨论的C的那个展开版本。

[13]参阅Hare（2），第36页。

[14]我们应当首先注意应该蕴含能够这个学说中"能够"的含义。假设有主张说，在某个情形中，我本应该以别的某个方式行动。按照该学说，只有当我本来能够这么做的时候，我才应该以这另一种方式行动。假如我本来就不能够以这另外一种方式行动的话，那么就不能够主张说，这是我本应该做的。正如我在第6节中所论证的，（1）这个主张——我本不能够以这另外一种方式行动——并不是（2）这个主张——鉴于我的实际愿望和意向，或简言之，鉴于我的动机，以这个方式行动本来就是不可能。主张毋宁是（3），就是，即便我的动机本来是不同的，以这个方式行动也原本是不可能的。无论我的动机原本是什么，以这个方式行动本来就是不可能的。

在主张应该蕴含能够方面我们有时候是正确的。假设有人认为：

（A） 未能拯救某人的生命总是错的。

假定我要么能够拯救一个生命，要么拯救另一个生命，但是我不能够两者都拯救。无论我拯救哪一个生命，我都会未能拯救某人的生命。根据（A），我不能避免错误地行动。只有当我拯救这两个人生命的时候，我才会避免错误地行动，而这是不可能的。凭借主张"应该蕴含能够"这个观点，我能够令人信服地拒斥（A）。我本来就不可能同时拯救两个生命。这不是主张说，鉴于我的实际动机这是不可能的。无论我的动机本会是什么，我都不可能拯救这两个人的生命。既然在这个意义上对于我而言两个生命都拯救是不可能的，那么就不能够辩解地主张说，这是我应该做的，两个生命未能都拯救，我是在错误地行动。

现在回到后果论。C主张说，对于任何人而言去做他相信将使结果更糟的事都是错误的。我们是在假定，倘若我们全都是行善者的话，那么结果就会比假如我们具有别的动机的情况下会有的更糟。如果我们相信这一点，那么我们决不做我们相信将使结果更糟的事就是不可能。我们将相信，如果我们是纯粹的行善者，那么我们就已经通过造成我们自己成为或者任由我们自己保持做纯粹的行善者而使结果更糟一些。如果我们不是纯粹的行善者，而是具有我们相信会使结果更好一些的动机，那么我们总是像纯粹的行善者那样行动——从不做我们认为将使结果更糟的事——是不可能的。按照这些选择方案中的任何一个，我们永不以这一方式行动都是不可能的。

从证明诉诸"应该蕴含能够"这一学说为正当合理的这个意义上讲，这是不可能的吗？无论我们的动机是什么或者本来可能是什么，我们决不以这个方式行动是不可能的吗？这是真的，但是具有误导性。它提示着，这个不可能性与我们的动机是什么没有什么干系。情况并非如此。这个不可能性实质上涉及一些有关我们动机的主张。我们决不做我们相信将使结果更糟的事这一点为什么是不可能的？这之所以不可能，是因为只存在一个意向，鉴于此意向，永不做我们相信将使结果更糟的事这一点会是因果上可能的，而造成我们自己具有或者保持这个意向这点本身（itself）就会是做我们相信将使结果更糟的事的一个事例。因为这个不可能性实质上涉及这些有关我们动机的主张，所以有关这是证明诉诸"应该蕴含能够"这个学说为正当方式的那类不可能性这一点并不清楚。但是至少能够说，这个事例非常不同于那个对于我来说两个人的生命都拯救并不可能的事例。那个不可能性与我的动机并不相关。在那个事例中，我们本能够令人信服地诉诸"应该蕴含能够"这个学说。这并不表明，我们能够令人信服地在这个与之非常不同的事例中诉诸这个学说。也许我们能够。但是我相信，我们在此不能够凭借主张说在那个我不能够两个生命都拯救的事例中诉诸这个学说是令人信服的这一点

来支持对这个学说的一种诉诸。因为它们是如此不同，我相信那个事例是不相关的。
该不相关性有可能遭到否定。如果我们相信假如我们全都是纯粹的行善者的话结果会更糟，那么我们永不做我们相信将使结果更糟的事是不可能的。我已经称这个主张——无论我们的动机是什么或者本来可能是什么这都是不可能的——是具有误导性的。这个主张虚假地提示着，这个不可能性与我们的动机无关。但是，纵使这个主张具有误导性，它仍然是真的。而且这或许可以主张为使这个事例与其中我不能够两个人的生命都拯救的那个事例充分地相似。
有没有这样的情形，其中即便当这种不可能与我们的动机无关的时候，我们能够否认应该蕴含能够？一些学者认为存在一些这样的情形。内格尔主张，可能存在一些道德悲剧，即这样一些情形，就是某人无论做什么他都将是在错误地行动。内格尔承认，在做出他的主张的时候，他是在否定"应该蕴含能够"。在他所描述的那些情形中，一个人无论动机如何，他都不能避免错误地行动，但是尽管如此，他也应该避免错误地行动。希望不会产生这样的一些情形是自然的。但是内格尔写道："……假设存在一个方法可以化解我们面对世界的每一个道德难题是天真的。"世界可能带给我们一些难题，对于它们我们没有什么避免做错的化解之道（Nagel（2），第79页）。
内格尔提到，之所以能够有这样一些情形，是因为那个最好的道德理论包含一些相冲突的原则。他可能主张："一个单一的道德原则不可能具有这样一些蕴含。如果这样的一个原则蕴含着我们不能够避免做错，那么这个原则就是不可辩护的。"当应用于像那个其中我不能够两个人都拯救的情况时，这个主张是令人信服的。原则（A）蕴含着，当未能使两个生命都得到拯救的时候，我是在错误地行动。此处之过在于这个原则，而不是在于这个世界。
假如倘若我们全都是行善者的话结果会更糟这一点为真，那么后果论蕴含着我们不能够避免错误地行动。尽管应该后果论者可能诉诸不同的道德原则，但是这个特定的结论却只是被一个单一的原则所蕴含。它是被这样一种主张所蕴含，那就是，做我们认为将使结果更糟的事是错的。但是，如果能够有一些其中我们不能够避免错误地行动的情形，正如内格尔所主张的，那么这个解释可能并非不得不是在两个不同的原则之间存在一种冲突。该解释不能够简单地诉诸这样一个事实，那就是，以这两种不同的方式都来行动是因果上不可能的。该解释必须诉诸某种更加深刻的东西——某种像两个不同的原则之间的冲突一样的东西。而且，在我们所考虑的那个事例中，该解释就可以被认定为属于这一种。存在一种冲突，但是不是在两个不同的原则之间，而是在最好的那组行动会是什么与最好的那组动机会是什么之间的冲突。我们是在假定，鉴于那些有关人的性质的事实，下面这样的一点是因果上不可能的：我们既实施会使结果最好的那组行动，又具有一旦具有便会使结果最好的那组动机。这种冲突可能被认为与两个不同的原则之间的冲突所产生出来的充分地相似。而且此处之过可以被认定为不是在于那个原则——做我们认为将使结果更糟的事是错的，而是在于这个世界。一个后果论者会重复内格尔的部分主张。他或许说："这样的一点可能是真的，那就是，如果我们具有将使我们能够永不做我们认为将使结果更糟的事的那个意向，那么我们就已经通过使我们自己具有或者保持这一个意向而使结果变得更糟了。如果这是真的，那么，世界就已经给予我们一个没有解决之道的难题。更确切一些说，尽管存在一种解决之道，但是它并不能够使我们避免做错。在把一切都考虑到的情况下，存在某种我们应该去做的事。我们应该使我

们自己具有或者保持我们一旦具有便将使结果更好的那些动机组之一。但是，如果我们的确具有这些动机组之一，那么我们永不做我们认为将使结果更糟的事这一点就是不可能的。如果在把一切都考虑到的情况下我们做我们应该做的，那么我们永不错误地行动这一点就是不可能的。”

在这个事例与内格尔所讨论的那些事例之间还有另一个差异。他提到，在某人不能够避免做错的那些事例中，他应当因其错误地行动而受到责备，而且应当感到懊悔和内疚。这是与应该蕴含能够这个学说冲突最为尖锐的那个主张。正如我说过的，最令人难以置信的是，某人无论做什么或者本来会做什么，他都该当受到责备，都应当感到懊悔和内疚。后果论并不蕴含这个主张。当克莱尔拯救她的孩子而不是那些陌生人的时候，她不应该受到什么责备，而且她不应当感到懊悔和内疚。大概她应当感到威廉斯所称的行动者之憾（agent-regret）。也许，当想到她本来能够拯救的、那些死去的陌生人的时候，她应当有这种感觉。但是这种感觉不是懊悔和内疚。

C并不蕴含着，克莱尔无论做什么或者本来会做什么，她都该受责备，而且应当感到懊悔和内疚。这就是威廉斯和内格尔有关面对道德悲剧的行动者所主张的。如果我们认为“应该蕴含能够”，我们可能拒斥这个主张。但这不会是拒斥C的一个理由，因为C并不蕴含这个主张。

如果一个后果论者拒斥这些说法，他就必须修正他的道德理论。假设他退一步承认，如果我们本来不可能总是避免做我们认为将使结果更糟的事，那么就不能够主张说这一点是我们总应该做的。不能够主张说，做我们认为将使结果更糟的事总是错误的。既然我们不可能永不以这个方式行动，那么以这个方式行动就不能够总是错的。必定有一些事例，其中尽管我们以这个方式行动，但并不是在错误地行动。根据（C3），做我们认为将使结果更糟的事总是错误的。一个后果论者现在或许放弃（C3）而代之以

（C3′） 如果以一种本身并不使结果更糟的方式行动这是可能的，那么我们总是应该避免做我们认为将使结果更糟的事。如果这是不可能的，每当我们本来可以以一种不会使结果更糟的方式不同地行动时，就应该避免做我们认为将使结果更糟的事。

正如我已经论证的，“不可能”在此并非意指“鉴于我们实际上的动机，因果上不可能”。它可能意味着“无论我们的动机本来会是什么，因果上的不可能”。假如我们全都是纯粹的行善者这一点如果会使结果更糟，那么我们永不做我们认为将使结果更糟的事就是因果上不可能的。无论我们的动机本来会是什么，这在因果上原本就是不可能的。既然我们永不以这种方式行动在这个意义上是因果上不可能的，那么（C3′）蕴含着，当我们以这个方式行动的时候，我们不能够总是在错误地行动。如果后果论主张（C3′）而不是（C3），那么它就呈现出一种修正了的形式，满足了“应该蕴含能够”这个学说。C的这个修正了的版本答复了诉诸这个学说的那个诘难。

如果C以这种方式得到修正的话，会带来多大差别？假设我们全都采纳（C3′），而且我们真的认为倘若我们全都是纯粹的行善者的话结果会更糟。假如我们具有别的许多可能的动机组之一的话，结果本来会更好。下面假定我们造成我们自己具有或者允许自己保持这些其他动机组之一是因果上可能的。（只有按照这个假定，对C的这个诘难才会出现。）（C3′）蕴含着我们应该造成我们自己具有或者保持这些别的动机组之一。既然以这些方式中的任何一个方式行动都是可能的，那么（C3′）蕴含着不去这么做会是错误的。

接着假定我们具有这些别的动机组之一。因为这是真的，我们时常做我们认为将使结果更糟的事。（C3′）蕴含着，每当我们本来可以以一种不会使结果更糟的方式来不同地行动的时候我们却以这种方式行动，我们就是在错误地行动。在所有这些情形中我们以一种原本不会使结果更糟的方式来不同地行动是可能的吗？我们在此不是在问，鉴于我们的实际动机这在因果上本来是否可能。我们是在问，假如我们的动机有所不同的话，这本来是否可能。对于这个问题的回答是否定的。在所有这些事例中，我们以一种不会使结果更糟的方式来不同地行动对于我们而言本来就是不可能的。只有当我们全都是纯粹的行善者的时候，我们才会在所有这些事例中不同地行动，而这就会使结果变得更糟。

我们现在应当问："我们本来能够在这些情形中的一些情形中以一种本来就不会使结果更糟的方式来不同地行动吗？"答案是肯定的。在不具有一个纯粹的行善者的意向的情况下我们总是像一个行善者一样行动是不可能的。但是，在不具有这个意向的情况下我们有时候以这个方式行动会是可能的。如果决定论不真，我们就会时常以这种方式行动了。那么（C3′）就会蕴含着，尽管我们以这个方式行动的时候并不总是在错误地行动，但我们经常是在错误地行动。通过以（C3′）来取代（C3），一个后果论者答复了诉诸应该蕴含能够这个学说的那个诘难。而这个修正并未带来多大差异。

[15]（此注释为1987年所加）我所举的例子有缺陷。对于我而言造成我自己失去这个意向并不会是错误的，如果我做到这一点是通过以我的敌人毁掉我的生涯为结果来拒绝帮助我的敌人的话。这个例子可以加以修补。（我们可能假设，他只会损毁我的生涯，因而能威胁做出进一步损害。）

[16]Sidgwick（1），第490页。

[17]Sen and Williams，第16页。

[18]"因为感到我得说出的最深刻的真理决不是'好音信'……如果我能够的话，我也不会……说出任何会使哲学——我的哲学——受欢迎的东西。"Sidwick（2），第395—396页。

[19]Williams，第145页；Sen and Williams，第15页。

[20]Williams，第145页；Sen and Williams，第15页。

[21]Williams，第145页；Sen and Williams，第15页。

[22]Williams，第145页；Sen and Williams，第15页。

[23]Rawls，第133和第182页。

[24]Williams（6），第3章。

[25]我把这个例子归功于T·斯坎伦。

[26]Newman，卷I，第204页。

第二章

实践上的两难

21.何以后果论不能是直接自败的

我已表述了何以有些理论是能够间接自败的。何以有些理论可能是直接自败的？比如说，当某个人成功地做了对他来说可能的那些行动中的那个最好地达到他的T设目标的行动时，他成功地遵循理论T。用我们意指“某群体的那些成员”。在以下情况下，我们可以称T

直接地集体性地自败的。倘使我们全都成功地遵循了T，我们从而将使我们的T设目标之达成比倘若我们中没有人成功地遵循T的话所会达成的更糟这一点为真。

这个定义看起来令人信服。“全都”和“没有”带给我们一些最简单的事例。讨论这些就足够了。

尽管看起来令人信服，但是我们必须拒斥这个定义。当它应用于某些协同问题时就不再令人信服了。这是这样一些事例，就是每一个人的行动后果取决于别人所做的是什么。第26节将涉及一个简单的事例。下面所示的是另一个事例。

		你 做(1)	你 做(2)	你 做(3)
我	做(1)	第三好	糟	糟
我	做(2)	糟	第二好	同等好
我	做(3)	糟	同等好	糟

如果我们两者都做（1），我们两者都成功地遵循C。既然你已经做了（1），那么如果我做（2）或（3）的话，我就会使结果更糟。而且你会提出相同的主张。如果代之以我们两者都做（2），那么就都没成功地遵循C。既然你已经做了（2），如果我做（3）的话，会使结果更好。而且你会提出相同的主张。如果我们两者都做（1）而

不是（2），那么我们两者都成功地而不是都没成功地遵循C。但是我们从而使结果更糟，使我们的C设目标之达成更糟。按照上述的定义，C在此是自败的。

这个结论并不是得到正当证明的。的确，如果我们两者都做（1），两者都成功地遵循C。但是，如果我们产生了最好的结果中的任何一个的话，我们本来也会成功地遵循C。如果我们两者中的一个做了（2），另一个做了（3），那么每一个就做了对他来说可能有的行动中会使结果最好的事。此处对C的诘难并非它是自败的。此处的诘难是，它是不确定的。无论我们两者都做（1），还是我们中的一个做（2）而另一个做（3），两种情况我们都成功地遵循C。由于这是真的，所以如果我们两者都成功地遵循C，那么并不确保我们的行动联合地产生最好的可能结果中的一个。但是也并不确保它们不产生。如果我们产生了最好的结果中的一个的话，我们本来就成功地遵循C了。C并不指引我们离开最好的结果。此处的诘难强度小一些。C只是未能指引我们走向这些结果。（我将在第26节中解释这个诘难如何能够是部分地得到答复的。[1]）

如果C指引我们离开那些最好的结果的话，那么如果我们成功地遵循C，我们将不产生最好的结果之一这一点会是肯定的。这提示着另一个定义。当出现以下两种情况时，称理论T是直接地集体性地自败的：

（i）当以下情况肯定时，如果我们都成功地遵循T，我们将从而使我们的T设目标之达成比我们没有人成功地遵循T的话所会达成的更糟，或

（ii）仅当我们并不成功遵循T的时候，我们的行动才会使我们的T设目标最好地达成。

情况（ii）表达的是这样一个观念，就是，要使我们的T设目标最好地达成，我们必须违背T。我用“当”并非意指“仅当”。我们无需涵盖所有的情形。正如我所解释的，“我们”并不意指“每一个活着的人”而是指“某群体中的所有成员”。

就C的事例而言（i）和（ii）会为真吗？仅当我们并不成功地遵循C的时候，我们才可能使结果最好这一点为真吗？因而肯定的会是，如果我们全都而不是没有人成功地遵循C，我们将从而使结果更糟吗？其中没有哪一个是可能的。当每一个人做出对他而言可能做出的那个使结果最好的行动时，我们成功地遵循了C。如果我们的行动的确共同地产生出最好的那个结果，那么我们必定都是在成功地遵循C。此处就任何人的情况而言，不可能为真的是，如果他以不同方式行动的话，他本来会使结果更好。按照这个定义，C不可能是直接自败的。

C可能是一个多元理论，通过诉诸数个不同的原则来评估结果的善。其一会是某个关于利弊的功利主义的主张；其他的或许是一些关于公正分配、关于欺骗、强迫或资格的原则。如果这些和其他原则告诉我们一致认同哪一个结果会更好的话，刚刚所作的推理将是适用的。这样的一个多元理论不可能是直接自败的，因为，它是对行动者中立的：给所有行动者只设定共同的道德目标。如果我们使这些共同目标最好地达到的话，我们必定是在成功地遵循这个理论。既然如此，那么只有倘使我们并不遵循这个理论时我们才将造成这些目标得到最好的实现这一点就不可能是真的。

22.有些理论如何会是直接自败的

如果我们的理论是相对于行动者的，给不同的行动者设定了不同的目标，又当如何？我们现在可能无法应用我的定义中的（ii）条。如果T给不同的人设定不同的目标，我们就没有什么能够最好地达到每一个人的T设目标的方式。但是我们略加修改却能够应用（i）条。而且我将给出另一个定义。

当如果某个人成功地遵循T，从而使他自己的T设目标的达成比他倘若没有成功地遵循T的话更糟这一点肯定时，我们称T是直接地个体性地自败的。

和

当如果我们全都成功遵循T，从而将使每一个人的T设目标的达成比我们当中没有人成功地遵循T的话所会达成的更糟这一点肯定时，我们称T是直接地集体性地自败的。

自利论为不同的行动者设定不同的目标。这个理论会是直接地个体性地自败的吗？S给我设定的目标是，我的生活对我来说要过得尽可能地好。当我做出我的可能行动中的那个对我而言将是最好的行动时，我成功地遵循了S。如果我成功遵循S，我将从而使结果对我更糟这一点会是肯定的吗？这不可能。无论就一个单一行动的情形而言，还是就不同时间的一系列行动的情况而言，都是不可能的。对这第二个主张的论证就像我上述所给出的那个论证。S给我设定了不同时间的同一个共同目标：我的生活对我来说尽可能过得好。如果我的不同时间的行动使我的生活尽可能过得好，我必定在每一个行动上成功地遵循了S。我必定是在做出我的可能行动中那个对我而言将是最好的行动。所以不能够肯定的是，如果我总是成功地遵循S，我将从而使结果对我而言更糟。

对我来说能够更糟的事是始终倾向于遵循S。但是在这个情形中，坏的不是我的行动而是我的意向。S不能够是直接地个体性地自败的。它只能够是间接地个体性地自败的。

有些理论能够是集体性地自败的吗？假设理论T给你和我设定了不同的目标。而且假设每一个要么会（1）促进他自己的T设目标，要么（2）更有效地促进另一个人的T设目标。结果表示如下。

		你：做（1）	你：做（2）
我	做（1）	每一个人的T设目标完成得第三好	我的目标完成得最好 你的目标完成得最糟
我	做（2）	我的目标完成得最糟 你的目标完成得最好	每一个人的T设目标完成得第二好

最后假设两者中没有哪一个选择将影响到另一个选择。那么就每一种情况而言这一点是真的，就是，如果他做（1）而不是（2），他将从而使他自己的T设目标得以更好地达到。无论另一个人做什么都是如此。只有倘若我们两者都做（1）而不是（2）的时候，我们两者才都成功地遵循T。只有那时每一个人才是在做其可能的行动中的那个最好地达到他的T设目标的事。但肯定的是，如果两者都成功地而不是两者都没成功地遵循T——如果两者都做（1）而不是（2）——我们从而将造成每一个的T设目标完成得更糟。理论T在此处是直接地集体性地自败的。

这样一些情形具有重要的实践意义。最简单的一些事例可能发生在以下情况中：

（a）当理论T是相对于行动者的，给不同的行动者设定不同的目标时，

（b）当每一个人的T设目标的达到部分取决于其他人做什么时，以及

（c）当每一个人所做的将不影响到其他人所做的时。

23.囚犯的两难困境和公共的善

如果T是自利论，那么这三个条件通常成立。S常常是直接地集体性地自败的。这些情形具有得自一个例子的一个误导性的名称。这就是囚犯的两难困境。这是一宗联合犯罪的案子，你和我分别受到审讯。结果表示如下。

无论另一个人做什么，对每一个人来说如果他认罪的话将更好。通过认罪，每一个人将肯定为自己减少两年监禁。但是如果两个人都认罪的话，对每一个人来说将比两个人倘若都保持沉默的话更糟。

		你	
		认罪	保持沉默
我	认罪	每人判10年徒刑	我获释 你判12年徒刑
	保持沉默	我判12年徒刑 你获释	每人判2年徒刑

让我们简化一下。如果每个人做而不是两个人都不做对他自己更好的事，对两人而言都将更糟。一种情形是，当

（肯定条件）每个人可能要么（1）使他自己得到两个利益中较小的那一个，要么会（2）给另一个人更大的利益，

以及

（否定条件）两个人的选择都不会以其他方式使两者中的任何一个更好或者更糟。

当肯定条件成立时，结果表示如下。

		你	
		做（1）	做（2）
我	做（1）	每个人得到 那个较小的利益	我两个利益都得到 你两个利益都没得到
	做（2）	我两个利益都没得到 你两个利益都得到	每个人得到 那个较大的利益

如果我们加上否定条件，那么表格变成下面这样。

		你	
		做（1）	做（2）
我	做（1）	对每个人来说，得到 第三好	对我来说最好， 对你来说最糟
	做（2）	对我来说最糟， 对你来说最好	对两者来说都是 第二好

否定条件有的部分不能够在这个表格中表现出来。一定是没有互惠：没有哪一个选择会造成另一个作出相同的选择这一点必定为真。如果他做（1）而不是（2），那么对每个人而言将更好。无论另一个人做什么都是如此。但是如果两者都做（1），这对每个人而言比倘若两人都做（2）更糟。

在什么时候两人中没有哪一个人的选择会影响到另一个人？仅当每一个人获悉另一个人所作选择之前必须作出最终选择的时候。在监狱之外，或者在博弈论专家的办公室之外，这很少为真。它也不会确保那个否定条件。可能有，例如，被延迟的互惠。两者中无论哪一个选择可能影响到他是否后来受害于或得益于另一个人。因此，我们很少能够知道我们面对一个囚犯的两难困境。

最后这一主张得到关于囚犯的两难困境的大量文献的支持。这类文献叙述了很少有说服力的两人事例。我的否定条件很难成立。

受到广泛讨论的那些事例之一是美国和苏联之间的军备竞赛。这时常被认为是一种囚犯的两难困境。这两个国家都应当秘密地发展新武器吗？如果只有一个国家这么做，它后来就可以摆布另一个。这将会是它的最好结果，另一个的最糟结果。如果两者都这么做，它们将保持同等地位，但是要付出相当大的代价，而且伴随着持续竞争的不安全感。这对两者会是第三好。对两者来说第二好是哪个国家都不秘密发展新武器。每一个都应该发展新武器，因为，如果另一个做同样的事的话，它将得到其第三好结果而不是其最糟结果；如果另一个并不做同样事的话，它将得到其最好结果而不是其第二好结果。但是如果两者都发展新武器，这比倘若两者都不发展新武器对两者来说都将更糟。[2]

我所说的否定条件有部分此处可能成立。如果新武器能够秘密地发展，每一国家在获悉另一个国家所作的选择之前必须作出它自己的选择。就研究问题而言，刚刚给出的推理可能是正确的。但是至于它是否适用于新武器的生产和部署，是否每一方知道对方在做什么，则

充满疑问。至于纯粹的研究进展是否能使一方摆布另一方也不清楚。而且，这还是一重复性的或持续性的状况。类似的决定必须一次又一次地作出。有鉴于此，以两种方式中的一种方式行事将肯定对每一方更好一些这一点不是很清楚的。每一方所作的选择可能影响到另一方后来的选择。

有大量的文献讨论这种重复性的事例：一种被误导性地称作重复性的囚犯的两难困境。为了弄清在这样一些事例中一对对的人如何行动，人们已经做了许多实验。[3]除了这样的一些实验性的工作之外，人们还对"重复性的囚犯的两难困境"进行了大量理论性的讨论。尽管有趣，但是这个讨论在此并不相干。我们应当区分两种事例。在第一种事例中，每一个人知道他将面对一定数目的"重复性的囚犯的两难困境"。鉴于就大多数具有实践重要性的事例而言这并不为真，我把对这些事例的讨论放到注释[4]中处理。

在绝大多数重要事例中，我们并不知道我们到底将多少次地面对"重复性的囚犯的两难困境"。按照我的定义，那些面对这样的一系列事例的人，甚至没有面对过哪怕是一个真正的囚犯的两难困境。就这些人而言这一点并不为真：就是如果每一个而不是两个都不做将对他自己更好一些的事，则将对两者会更糟。这一点之所以并不为真，是因为在这些"重复性的囚犯的两难困境"中，两个选择中究竟哪一个对自身将更好不再清楚。这是因为一个人的选择可能影响到他人后来的选择。如果一个人作出协作性的选择，这可能引导他人后来作出相同的选择。正如博弈论专家所说的，如果我们考虑到选择的所有可能后果，就没有哪一个是主导性的，就没有哪一个肯定是对自身而言更好的。因而，这样的一些事例所产生出的问题对自利论者而言是一个内部的问题。如果一个人的目标是尽其所能地做那对自身最好的事，一个人在"重复性的囚犯的两难困境"系列中应当如何行动？在一个真正的囚犯的两难困境中，所产生出的问题则有所不同。在一个真正的两难中，如果一个人以两种方式中的一种方式行动，那么肯定的是，这

对他自身将更好一些，不仅当下如此，而且从长远来说和从总体上来说也是如此。所产生的问题不是一个人如何能够最好地追求自己利益的那个内部的问题。所产生的问题是，如果每一个人而不是两个都不做那个对自己肯定更好的事，那么对他们两者而言都将更糟。

尽管我们很少能够知道我们面对着一个囚犯的两难困境，但是我们经常能够知道我们面对着那些多人版本。而且这些具有重大的实践上的重要性。罕见的两人事例只有作为多人版本的一个模型才是重要的。

如果每一个而不是我们两个都不做对自己更好的事，这对每一个人将会更糟这一点肯定时，我们面对着一个多人的两难困境。

这个定义只涵盖那些最简单的事例。如前所说，“每一个人”意指“某个群体中的所有的人”。

多人情形即撒玛利亚人的两难[5]。我们每一个人有时在自身付出较少代价的时候会帮助一个陌生人。每一个人也会大约同样地经常得到帮助。在一些小共同体中，助人所付出的代价或许间接地得到偿还。如果我助人的话，作为回报，这可能导致我后来得到帮助。但是在大共同体中，则未必如此。此处如果每一个人从不助人可能对他更好。但是，如果没有什么人助人，则对每一个人将会更糟。每一个人或许因从不助人而得益，但是他会因从未得到帮助而受损，而且受损得更多。

很多事例发生，当

（肯定条件）（i）我们每一个人在自己付出一定代价的情况下给予别人一个更大的利益总量或更大的预期利益总量；（ii）如果每一个人而不是没有人把这个更大的利益给予别人，每一个人会得到一个更大的利益或者一个更大的预期利益时；以及

（否定条件）不存在抵消掉这些直接后果的间接后果时。

肯定条件涵盖很多种情形。在一个极端，每一个人都会给那些别人中的一个人以更大的利益。这方面的一个例子就是撒玛利亚人的两难。在另一个极端，每一个会给那些别人中的所有人一个更大的利益总量。在这两个极端之间的一些事例中，每一个会给那些别人中的一些人以更大的利益总量。（在第二个极端，其中每一个会有益于所有那些别人，（ii）是多余的，因为它可以被（i）蕴含。在其他事例中，（ii）时常为真。例如，如果利益是随机分布的，那么它就会为真。）

另一事例的范围牵涉到每一个人所做的会有益于别人的几率。在一个极端，每一个人肯定会给予别人更大的一个利益总量。在另一个极端，每一个人给予别人一个更大许多的利益是一个非常小的几率。在这个事例范围中，每一个人会给予别人一个更大的预期的利益总量。这就是那些可能的利益乘以该行动产生这些可能利益的几率后所得到的值。当我们的行动后果不肯定的时候，我有关那个两难的定义就需要予以修正。在这些事例中，这一点是不肯定的，就是，如果每一个而不是大家都不做将对自己更好一些的事，这对每一个人将会更糟。

当如果每一个而不是大家都不带给自己一个预期的利益，这要么将减少对每一个人来说的预期利益，要么将施加在每一个人身上以一个预期的损害或代价肯定时，我们面对着一个风险两难的困境。

在一些多人事例中，只有肯定条件成立。在这些事例中，因为所牵涉到的那些人数足够地少，每一个人所做的可能影响到别的大多数人所做的。这些事例具有实践上的重要性。有许多事例牵涉到国家、公司或工会。这些事例具有一个真正的囚犯的两难困境的一些特征。但是它们缺乏那个中心特征。因为每一个人的行动可能影响到足够多的那些他人的行动，所以哪一个行动会符合每一个人的利益是不清楚的。这样的一些事例所产生出的问题又是一个对自利论者而言的内部

的问题。在一个真正的囚犯的两难困境中，至于哪一个行动将总体上说来带给行动者以一个更大的利益或者一个更大的预期利益，不存在什么不确定性。真正的那些两难困境所产生出的问题是相当不同的。

我说过，多人两难是极为常见的。理由是这样的。在一个两人事例中，否定条件未必成立。这可能需要狱警或者博弈论专家予以特别确保。但是在那些牵涉到许多人的事例中，否定条件自然成立。每一个人必须在获悉别人做什么之前行动这一点无需为真。即便当这一点不为真的时候，如果我们人数非常庞大，每一个人所做的也极不大可能影响到绝大多数他人所做的。它可能影响到一些他人所做的；但是这很少会造成足够的不同之处。

最常见的真正两难是那些贡献者两难。这些牵涉到公共的善：后果甚至有益于那些并不帮助产生这些后果的人们。就每一个人的情况而言，这一点能够是真的，就是如果他帮助，他将增加利益总量或预期的利益量。但是他所增加的利益中只有非常小的一部分会返回到他。既然他分享的只是他所增加的东西中的一个非常小的份额，它也许不能补偿他的贡献。于是，对每个人说来，如果他不做贡献对他可能更好。无论他人做什么这都能够如此。但是如果做贡献的他人更少，则对每一个人说来将会更糟。而且如果没有人做贡献，这将比如果所有的人做贡献更糟。

很多贡献者两难牵涉到两个临界值。在这些事例中，存在v和w这样两个数值，如果做贡献的人数比v少，那么就没有什么利益产生出来，而且如果做贡献的人数比w多，这将不会增加所产生的利益。在许多这样的事例中，我们并不知道他人可能会做什么。那么，这一点将是不肯定的：就是如果有任何人做贡献的话，他将有益于他人。只有他带给他人一个预期的利益时这点才将是真的。一个极端的事例是投票，其中两个临界值之间可能只有一票之差。数值w此处等于$v+1$。尽管选举很少是一个真正的囚犯的两难困境，但还是值得留待更后面的相关地方予以讨论。

一些公共的善需要财政贡献。就道路、警务和国防而言都是如此。其他一些则需要合作努力。当大型企业中工资取决于利润，而且工作是不快乐的或者是一种负担的时候，如果他人工作更努力一些，那么对每一个人会更好一些；如果他自己工作更努力一些，那么对每一个人会更糟。就集体农场的农民而言同样如此。第三种公共的善是避恶。此处所需要的贡献通常是自制。这样的一些事例可能涉及

上班族：如果他驾车的话，对个人来说会走得更快，但是如果全都驾车的话，每一个人则比全都乘坐公共汽车的话走得更慢；

士兵：如果他转身逃跑，对他而言将更安全一些，但是如果全都这样的话，被杀死的人将比大家都不这样做更多；

渔民：当海洋过度捕捞时，如果他设法捕捞更多的话，对他而言可能更好一些，但如果渔民全都这么做的话，对每一个人而言则更糟；

农民：当土地过度拥挤的时候，如果他或她生育更多的孩子的话，对他而言能够更好一些，但如果大家全都这么做的话，对每一个人来说则会更糟。[6]

还有无数其他事例。就每一个人而言，如果他增加污染、消耗更多的能量、插队和撕毁协议，能够更好一些；但是如果全都这么做的话，对每一个人来说则比大家都不这么做更糟。如果每一个人而不是大家都不做那些对他自己而言将更好的事，则对每一个人来说都将更糟这一点通常都是真的。

这些两难困境都是以利己的标准表述的。既然鲜有人是纯粹利己的，这可能看起来减少了这些事例的重要性。但是就这些事例中的绝大多数而言，下面这一点是真的。如果每一个人而不是大家都不做将对他自己，或他的家庭，或他所爱的那些人更好一些的事，那么对每一个人来说则将更糟。

24.实践上的难题及其解决方法

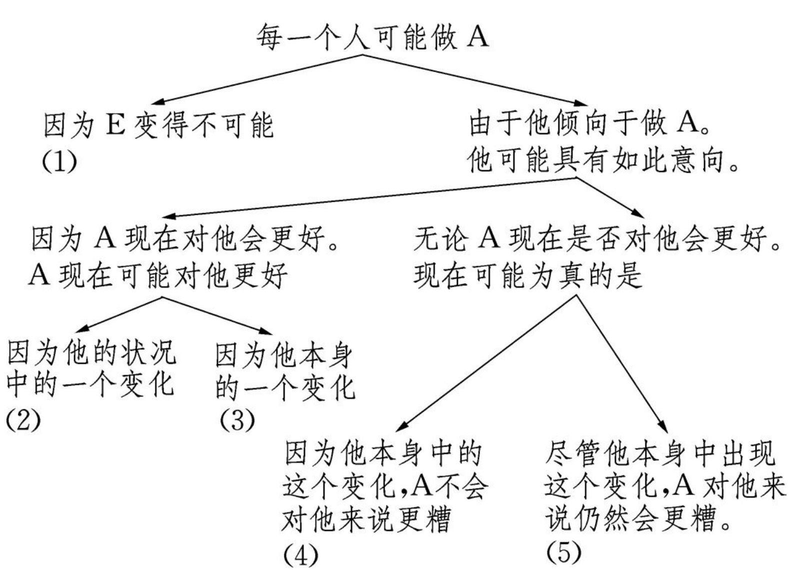
假设每一个人都倾向于做将对他自己或者他的家庭抑或他所爱的人们更好的事。那么就有一个实践上的难题。除非出现某种变化，否

则实际的结果对每一个人都将会更糟。这个难题是我们为什么需要比自由竞争经济得多的东西的主要理由之一，即我们为什么需要政治和道德的主要理由之一。

且让我们使用贴标签的方法。而且采用人们所理解的"或他们的家庭抑或他所爱的人"这些语汇。每一个人具有两个选择余地：E（更利己一些）和A（更利他一些）。如果全都做E，那么这比全都做A对每一个人更糟。但是无论他人做什么，就每一个人而言如果他做E将更好一些。该难题就在于，因为这个理由，每一个人现在都倾向于做E。

如果绝大多数人做A，这个难题将部分得到解决，如果全都做A该难题将完全得到解决。以下面图示中所表明的方式中的一个方式或多个方式可能得到一个解决方法。

（3）中的变化不同于（4）中的变化。在（4）中某人倾向于做A，无论如何这对他而言更好。因为他本身中的这个变化，A对他来说不会



更糟这一点是一个纯粹的副后果。在（3）中某个人倾向于做A，只是因为，鉴于他本身的某个别的变化，做A对他会更好一些。

（1）到（4）消除了那个两难困境。利他主义的选择对每一个人不再会是更糟的了。这些通常是好的解决方法。但是它们有时是无效率的，或者是难达到的。那么我们需要（5）。（5）解决那个实践上的难题，但是它并不消除那个两难问题。一个理论性的难题仍然存在。在本章和下一章中我讨论我们如何能够解决那个实践上的难题。我在第四章中讨论那个理论性的难题。

在解决方法（1）中，自利的选择是不可能的。这有时是最好的。在很多贡献者两难中，应当存在不能逃避的纳税。但是（1）经常会是一种拙劣的解决方法。渔网会被撕毁，士兵会被绑在他们的哨位上。两者都具有不利之处。

（2）是一种较不直接的解决方法。E仍然是可能的，但是对每一个人来说A更好。或许有一套奖励制度。但是，如果它起作用的话，所有的人都必须得到奖赏。假使惟一的奖励是避免某种惩罚，可能更好。如果这起作用，就没有什么人遭受惩罚。如果所有该当惩罚者都被枪毙，可能就没有什么该当惩罚者。

（1）和（2）之间的选择常常是困难的。考虑一下那个农民的两难。其中就每一个人而言如果生育更多孩子的话，能够更好，如果全都这么做的话，对每一个人来说则更糟。一些国家奖励两孩家庭。中国现今奖励一孩家庭。但是在该难题最严重的地方，国家太穷，无力给每一个人奖励。而且如果这样一个制度要生效的话，非奖励必须像惩罚一样。由于该制度不会是完全生效的，所以某些人就会不得不承受这样的惩罚。这样的惩罚不仅落到父母头上，而且落到那些孩子们头上。

一个可供选择的余地是（1），其中对人们来说生育两个以上的孩子是不可能的。这会涉及在生下第二个孩子之后强迫绝育。万一有哪一个孩子夭折或者两个孩子都夭折，如果这个绝育是可以扭转的话，则会更好。这样的一种解决方法可能看起来有些吓人。但是在公民投票中它或许得到一致的支持。如果大家都不是全部生育两个以上

的子女，对某个群体中所有的人来说会更好。如果所有人都宁愿事情这样发生，那么所有的人可能选择并且投票支持这样一种强制性的绝育。如果它在公民投票中得到一致的支持，这或许会消除强制性中的吓人之处。而且这种解决方法比奖惩制度具有优势。正如我说过的，当这样的一种制度不是完全生效的话，那些生育更多孩子的人以及他们的孩子都必定受到惩罚。如果代之使会受惩罚的东西成为不可能，可能更好。[7]

（1）和（2）是政治的解决方法。被改变的是我们的状况。（3）到（5）是心理的解决方法。变化的正是我们。这种变化可能是具体的，只解决一个两难问题。渔民们或许变得懒惰起来，士兵们或许变得宁死不屈，或者被训练得自动服从。下面是四个更具一般性的那类变化：

我们可能变得值得信赖。那么每一个人可能答应去做A，条件是别人作出相同的承诺。

我们可能变得不愿做"搭便车者"[8]。如果每一个人认为别的许多人将做A的话，那么他可能宁愿做他的份额。

我们可能变成康德主义者。那么每一个人只会做他合理地意欲每一个人去做的。没有人会合理地意欲人们全都做E。因而每一个人会做A。

我们可能变得更加利他。鉴于充分多的利他主义，每一个人会做A。

这些是道德上的解决方法。因为它们或许解决任何一个两难问题，它们是最重要的心理解决方法。

它们通常比政治上的解决方法要好一些。这部分地是因为它们无需被强迫。就拿"撒玛利亚人的两难"为例。不能使不帮陌生人这一点为不可能。那些坏撒玛利亚人不能被轻易抓住和受罚。那些好撒玛利亚人会得到奖赏。但要确保这一点，法律就不得不介入。鉴于执行的代价，这种解决方法可能并不值得。如果我们直接倾向于帮助陌生人，会好得多。

知道哪一种解决方法会是最好的还不够。任何一种解决方法都必须得到实现或者贯彻。就政治解决方法而言这通常更容易一些。状况

比人更容易改变。但是我们时常面对另一个、二阶的贡献者两难。极少有政治解决方法能够由单一的个人来实现。绝大多数情况下需要很多人的协作。但是一种解决方法是一种公共的善，带给每一个人利益，无论他在贯彻这种解决方法方面是否尽了其份额。在绝大多数大型群体中，就每一个人而言如果他做了他的份额则将会更糟。他所造成的差异将小得无法偿付他的贡献。

这个难题在组织良好的民主政体中可能是小的。此处可能足以使原初的那个两难困境得到广泛的理解。这可能是困难的。但是我们从而可能为一种政治解决方法进行投票。如果我们的政府对民意调查有所反应，甚至可能无需进行投票。

当没有什么政府的时候，该难题就大得多了。这正是当年令霍布斯感到担忧之处。现在它应当是令各个国家感到担忧之处。例子之一是核扩散问题。在没有世界政府的情况下，获得一种解决方法可能是艰难的。[9]

当其解决方法受到某些统治集团反对的时候，该难题就再大不过了。这就是受压迫者两难。

这样的贡献者两难时常需要道德上的解决方法。我们时常需要一些人直接倾向于做他们的份额。如果这些人能够改变这个状况，以便得到一种政治解决方法，那么这个解决方法可能是自我支持的。但是没有这样的一些人，它可能永远不可能实现。

那么道德上的解决方法时常是最好的；而且它们时常是惟一的一些可以达到的解决方法。我们因而需要一些道德动机。如何能够引发这些动机？幸运的是，这不是我们的难题。它们存在着。这就是我们如何得以解决许多囚犯的两难困境。我们所需要的是使这些动机更加强烈，分布更加广泛。

就这个任务而言，理论大显身手。那些囚犯的两难困境需要得到解释。它们的道德解决方法也是如此。两者为人们所理解得都太少了。

我们看到，有种解决方法是一个条件协议。这要成为可能，我们都能够交流这一点首先必须为真。如果我们是纯粹利己的或者是从不克己的，交流能力很少会有差异。在绝大多数大型群体中，允诺我们将作出利他主义的选择这一点会是毫无意义的，因为就每一个人而言如果他遵守诺言的话对他会更糟。但是假设我们是值得信赖的。每一个人现在能够允诺去做A，条件是其他每一个人作出相同的承诺。如果我们知道我们都是值得信赖的，那么每一个人将有一个加入这个条件协议的动机。每一个人将知道，除非他加入，否则该协议就不会生效。一旦我们全都作出这个承诺，我们就将全都做A。

如果我们的人数是庞大的，全体一致在实践上将是难以达到的。如果我们惟一的道德动机是值得信赖的，那么我们将未必达成联合的条件协议。如果他加入的话，对每一个人来说很可能会更糟。（我们甚至也未必将交流。）[10]

很少有人的惟一的道德动机是值得信赖的。假设我们也不愿做搭便车者。如果我们每一个人具有这个动机，他将不希望留在那个联合的条件协议之外。他将宁愿加入，即便这样做对他将会更糟。这为该联合协议解决了刚刚提到的那个难题。而且，如果有足够多的人们不愿做搭便车者，将不会有达成一个实际协议的需要。所需要的一切无非是确保将有许多人做A。那么每一个人会宁愿做他的份额。但是不愿做搭便车者本身并不能够提供这个保证。所以存在许多它并不提供解决方法的事例。[11]

康德的检验标准总会提供一种解决方法。这个检验标准具有其自身的一些问题。我能够合理地要么希望没有人行医，要么希望所有人都行医吗？如果我们对这个检验标准加以提炼，我们也许能够解决这样的一些问题。但是在囚犯的两难困境中，它们并不出现。这些就是其中我们自然地说“如果每一个人都那样做的话会怎样？”的情形。[12]

第四种道德上的解决方法是充分利他主义的。我在此不是指纯粹的利他主义。纯粹的利他主义者，不给他们自己的利益什么权重，可

能面对类似囚犯的两难困境的情况。能够为真的是，如果所有人而不是没有人做肯定对他人更好的事，这对每一个人来说将会更糟。[13]至于“充分的利他主义”，我意指充分关注别人，限制的情形是不偏不倚的仁爱：对每一个人，包括自己在内，同等关心。

第四种解决方法是最不为人理解的。时不时地有这样的主张，就是在那些牵涉到非常多的人的贡献者的两难困境中，每个人所做的不会有什么差异。这一直被认为表明合理的利他主义者不会有什么贡献。正如我将在下一章中论述的，这并非如此。

---

[1]在这些段落中我只是步里根之后尘。

[2]这一点，例如，在Gauthier（1）和Brams中有论证。

[3]在这些实验中，那两个选择被称作是合作性的和非合作性的。例如，业已得到表明的是，如果那些参与者服下5毫克的安定，他们更有可能合作。参见《冲突之解决杂志》诸期，安·阿伯，密歇根。

[4]常有论证说，如果各人知道他将面对一个特定数目的“重复的囚犯的两难困境”，那么，对于各人而言，如果他总是作出那个非合作的选择将更好一些。这个选择对最后一轮中的各人将更好。既然各人知道这一点，那么这个选择对于倒数第二轮中的各人将更好。相似的推理回溯到第一轮。

[5]撒玛利亚人是古代的一个民族。在希伯来人的北国以色列亡于亚述（公元前722年）之前，此名是指定都撒玛利亚的北国以色列人，之后则指亚述王从外地迁来的居民与当地居民杂居而形成的新民族。一般认为，后世撒玛利亚人有利可图时就承认自己与犹太人有血缘联系，引火烧身时就矢口否认这种联系。与该两难相关的内容参见《圣经·新约》中耶稣所讲述的那个帮助陌生人的“好撒玛利亚人”典故（《路加福音》（第10章）。——译者

[6]约翰·布鲁姆对此提出质疑，他评论说：“如果一个额外的孩子所生产出的东西比他所消耗的要多，那么他的存在对于每一个人而言更好；而如果他消耗的比他所生产出的东西要多，那么他的存在对于包括他自己的父母在内的每一个人而言则更糟。”我相信这个诘难已经被老年多子多女所带来的那个大得多的好处所答复。

[7]参看Sen and Runciman。

[8]搭便车者（free-rider）：原意指搭乘顺风车的人，引申为“免费享用公共利益者”，“无功受禄者”。——译者

[9]核扩散问题的这个版本在某个方面其尖锐性更小，因为行动者的数目相对来说是小的。各人所做的可能影响到他人所做的这一点在此是可能的。

[10]假设我们所要求的某物要比全体一致要少一些。在一个有1000个成员的共同体中，每人可能以至少别的900人参加为条件承诺做A。如果我们惟一的道德动机是值得信赖，那么我们将不会达到这个解决办法。如果他作出这个承诺，对于各人而言很有可能会更糟。有三

种可能性。加入的其他人的数目可能最终（a）少于900、（b）正好900或者（3）多于900。如果（a）发生，则没有什么人的承诺具有约束性；所以当决定什么符合我们利益的时候我们可以忽略这个可能性。如果（b）发生，对于作出承诺的各人而言将更好。就各自而言这样的一点再次为真，就是，在没有他的承诺的情况下，整个计划会失败——没有别的什么人的承诺会变得具有约束性。对各人而言那会更糟。但是如果（c）发生，对于各个做出了他的承诺的人而言将更糟。这对于各个现在承诺做出那个利他性选择的人而言也是真的。既然他是值得信赖的，他将作出这个选择。但是即便没有他的承诺，别的每一个人的承诺也会变得具有约束性。所以各人将作出对于他自己而言更糟的选择，而且他将得不到什么回报。各人从他人的那些承诺中所得到的，即便在他本人没有作出该承诺的情况下，他本来也会得到。下面请注意，与（b）相比，（c）更加有可能发生。多于900个别的人参加比正好900个别的人参加的可能性要大得多。所以如果各人作出这个承诺，这对于他而言很有可能更糟。如果人数更多一些，比方说数以百万计，甚至更有可能如此。要么各人的承诺将不造成什么影响，因为太少的他人参加；要么对于他而言将更糟一些，因为多于足够数目的人参加。在其中作出该承诺最终将对于各人更好的那个情形，即在其中正好特定数目的人参加的那个情形，将是最不大可能的。可能有某种避免这个难题的方法，或许以一系列会聚于一点的"民意调查"。但是在那些牵涉庞大人数的情形中，则存在另一个难题。要能够使所有的人交流，然后达到一个联合的协议，会花费某些努力。但是该协议是一种公共的善，造福于各人，无论他是否促成这个协议。对于这个贡献者两难，纯粹的值得信赖这一点是提供不出什么解决之道的。

[11]参见Sen（3）。

[12]参见，例如，Strang and Ullmann-Margalit。

[13]在一个两人事例中，假设各个纯粹的利他主义者要么会（1）给予他自己一个更大的利益，要么（2）给予他人一个更小的利益。纯粹的利他主义在此对于两者而言都会更糟。（两个利益之间的差别可能得是巨大的，因为符合我们自己的利益的东西部分地依赖于我们的动机是什么。）

第三章

道德数学中的五大错误

常常有人认为，在那些包含非常多的人的事例中，任何单一的利他选择都不会有任何差异。持这一主张的人中有一些人相信，它只是从根本上削弱了由充分利他主义所提供的那第四个道德的解决方法。这些人论证说，在这样一些事例中，因为他们不能够诉诸我们行动的后果，所以我们必须代之以要么诉诸康德的检验标准——"如果每一个人都那样做的话会怎样？"，要么诉诸不愿搭便车这一点[1]。但是，如果我的贡献牵涉到我付出真正的代价，而且肯定不会带来什么差异——不会给他人带来什么利益，那么我可能不为我的不愿搭便车所动。只有当我相信我是在以牺牲他人为代价而获益的时候，这种不愿意才可能适用。如果我的贡献不会带来什么差异，那么我未能作出贡献这一点对他人而言将不会更糟，所以我将获得的利益并不是以牺牲他人为代价。我可能相信，该事例就像显然已经跨过某个临界值的那些事例一样，以至于任何进一步的利他主义行动纯粹是在浪费努力。这个信念也可能从根本上削弱那个康德主义的解决方法。如果我的贡献不会带来什么差异，我就能够合理地希望其他每一个人都按我所做的那样而行动。当知道其贡献不会带来什么差异的时候，我能够合理地希望，没有哪一个人作出贡献。既然其他人可能像我一样思考，那么，在这些事例中，我们凭借诉诸我们行动的后果是否能够解释我们为什么应当做出贡献，就有极大的重要性。

25.总体份额的观点

我们能够解释。但是要做到这一点我们必须避免道德数学中的几个错误。考虑一下

第一救援队：我知道下面的一切。100名矿工被水困在矿井中。这些人能够被升降机带上地面，而升降机是借助长长的杠杆上的重物来提升的。如果我与另外3个人站到某个平台上，重量刚好够提升起升降机，就救了100个人的性命。如果我不加入这个救援队，我就能够到别处去，单枪匹马地救出其他10个人的性命。有第五个潜在的救援者。如果我到别处去，这个人将加入到那3个人之中，他们4个人一道救出那100名矿工。

当我本可以有数种方式行动的时候，我应当如何决定哪一个行动给人们带来的利益最大？首先假设我们5个人都去救那些矿工。按照总体份额的观点，每一个人都产生出总体利益中他自己的份额。既然我们5个人救100个人的性命，那么每一个人救了20个人的性命。用不那么拘泥于字面的说法，就是每一个人所行的善等于所救的那么多的人。按照这个观点，我应当加入到另外那4个人之中，救出相等于20个人的性命。我不应当到别处去救别的10个人的性命，因为那样我所救的人数就会少一些。这显然是个错误答案。如果我与其他4个人一道的话，那10个人就会不必要地死去。既然其他4个人在没有我帮助的情况下也能救出那100名矿工，那么我应当去救这10个人的性命。

总体份额的观点或许可以修正。或许可以主张说，当我加入那些正在行善的他人之中的时候，我所行之善并非就是所产生出的总体利益中的我的份额。我应当从我的份额中减去我的加入所造成的他人的利益份额的缩减量。如果我加入这个救援队，我将是一起救出100个人的那5个人中的1个人。我的份额将是20个人。如果我没有加入的话，其他4个人就会救出那100个人，每一个人所救出的人的份额是25个人，或者说与我参加的情况相比份额多出5个人。通过加入其中，我减低了其他4个人所救的人数的份额，缩减的人数是5的4倍，或者说20个人。按照这个修正了的观点，我的利益份额因而是20减去20，或者说是0。因此我应当去救别的10个人。这个修正了的观点给出了正确答案。

下面考虑

第二救援队：像前面一样，100个人的性命处于危险之中。如果我和其他3个人加入一个救援队，这些人就能够得救。只有我们4个人能够加入这个救援队。如果我们中任何一个未能加入的话，那100个人都会死掉。如果我不加入，我可以到别处单枪匹马地救出50个别的人。

按照修正的总体份额观点，我应该到别处去救这50个人。相反如果我加入这个救援队的话，所产生的利益中我所分摊的份额只是等于救25个人的性命。从而，如果我到别处救50个人的话，我所为之善更多一些。这显然为假，因为如果我以这种方式行事的话，多于50个人的生命将会失去。我应该加入这个救援队。我们必须作进一步的修正。我必须在所产生的利益中我所分摊的份额之上加上我所导致的他人所产生的份额方面的任何增加量。如果我加入，我能够使那3个人中的每一个人与我一道救100个人。如果我不加入，这3个人将救不了任何一条命。我所分摊的份额是25个人，而且我把其他人所产生的份额增加了75个人。按照这个再次修正了的观点，我的总份额是100个人。这也是其他每一个人所产生的总份额。既然每一个人算作产生了这个总体利益之全部，那么这并非是总体份额观点的一个版本。它是一个与之完全不同的观点。这个再度修正了的观点给出了这个事例的正确答案。它并不是对宣称每一个人救100个人的性命这种观点的什么诘难。这是每一个人所做的，尽管并非是他单独做的，但确实是在他人的帮助下他所做的。[2]

这个观点可以更加简明地予以表达。我应当以其后果是最多的人得救的那个方式行动。更宽泛一些说，

（C6）如果一个行动的后果是某个人得益更多一些，那么这个行动就有益于某个人。如果一个行动的后果是某个人受损更多，那么这个行动就有害于某个人。最有益于人们的行动是那个其后果是人们受益最大的行动。

这些主张正确地蕴含着，我不应当加入第一救援队，而是应当加入第二救援队。

后果论者应当诉诸（C6）。如果其他人给予罗尔斯所称的善行原则任何权重的话，也应当如此。按照任何令人信服的道德理论，我们有时应当力求做会带给人们最大益处的事。

（C6）可能需要加以进一步解释。假设我要么能够做（1）要么能够做（2）。在决定哪一个带给人们益处更多的时候，我应当比较如果我做（1）的话人们后来会得到的所有益处和损失，以及比较如果我做（2）的话人们后来会得到的所有益处和损失。带给人们更多利益的那个行动是这样一个行动，在这个比较中会伴随着更大的净利益总量——减去损失之后的更大利益总量。如果，如经常为真的那些情形一样，许多其他人的行动也会是造成得到这些利益或损失的原因中的一些部分，也并不相干。

（C6）修正了"利益"和"损害"这些词语的通常用法。当我宣称已经给某个人带来利益的时候，我通常被理解为我的某个行动是这个人所得到的这个利益的主要原因或者直接原因。按照（C6），即便当我的行动是造成得到这个利益的原因中的一个遥远的部分的时候，我也是给这个人带来利益。所需要为真的是，如果我以其他方式行动的话，这个人本来不会得到这个利益。同样的主张适用于"损害"。

（C6）还在另一个方面修正了我们对"利益"和"损害"的用法。按照通常用法，即便有时我正在做的并非对这个人而言更好一些，我也是在给某个人带来利益。当我的行动，尽管足以产生某个利益，但是并非必需的时候，情况就是如此。假设我能够轻易地要么拯救J的性命，要么保住K的手臂。我知道，如果我不救J的性命，某个别的人肯定会救J；但是没有别的什么人能够保住K的手臂。按照我们的通常用法，如果我救J的性命，我给他带来利益，而且我带给他的利益要大于我保住K的手臂所带给K的利益。但是，出于道德目的，这不是度量利益的方法。在作出我的决定的时候，我应当忽略对J的这个利益，这是（C6）所告诉我要做的。根据（C6），当我救J的性命的时候，我并不给他带来利益。我的行动的后果是J得到更多的利益这一点并不为

真。如果我的行动有所不同的话，别的某个人会救J的性命。（C6）正确地蕴含着我应当保住K的手臂。这就是那种其后果是人们得到更多利益的行动。按照“利益”的已修正的用法，这是那个给人们带来更多利益的行动。

道德数学中的第一个错误是总体份额的观点。我们应当拒斥这个观点，代之以诉诸（C6）。

26.忽视行动集合的后果

人们自然地假定

（第二个错误）如果某个行动的对错是因为它的后果，那么惟一相关的后果就是这个特定的行动的后果。

这个假定至少在两类事例中是错误的。

事例一。X和Y同时开枪要杀死我。两者中无论哪一个开枪本身都会杀死我。

X或者Y都不是以一种后果是额外一个人死亡的方式行动。鉴于另一个人所做的，对每一个人来说这一点为真，那就是如果他没有对我开枪的话，本来也不会造成什么不同结果。根据（C6），无论是X或是Y都没有伤害我。假设我们犯了第二个错误：我们假定，如果说某个行动是错的是因为它的后果，那么惟一相关的后果就是这个特定的行动的后果。既然无论X或是Y都没有伤害我，我们就被迫得出这样一个荒谬的结论，那就是这两个杀人犯的行动没有错。

一些人会用这个事例来表明我们应当拒斥（C6）。还有一个更好的可供选择的余地。我们应当增加考虑：

（C7）纵使一个行动并不伤害任何一个人，这个行动也可能是错的，因为它是一个合起来伤害到他人的行动集合中的一个。同样，即便某个行动并不给人带来利益，它也可能是某

个人所应该做的，因为它是一个合起来给他人带来利益的行动集合中的一个。

X和Y的行动是错误的，因为他们合起来伤害到我。他们合起来杀死了我。（C7）甚至应当为非后果论者所接受。按照任何可信的道德理论，在这样一个事例中只考虑单一行动的后果是一个错误。按照任何令人信服的理论，纵使我们每一个并不伤害到什么人，但如果我们合起来伤害到别人，我们也可能是在错误地行动。

在事例一中，多因素决定的行动是同时发生的。在情况并非如此的那些事例中我们应当主张什么？考虑一下

事例二。X哄骗我喝下毒药，是几分钟内就造成痛苦死亡的那种。赶在这种毒药起作用之前，Y无痛苦地杀死了我。

尽管Y杀死了我，但是Y的行动对我并非更糟。（C6）因而蕴含着，在杀死我方面Y并不伤害到我。（Y的行动在一个方面对我而言略微更糟一些，因为它把我的生命缩短了几分钟。但此点有这样一个压倒性的事实，就是Y使我避免了痛苦地死亡。）（C6）还蕴含着，X并不伤害我。就像在事例一中那样，X与Y都不伤害我。（C7）正确地蕴含着，X和Y行动错误，因为他们合起来伤害到我。他们之所以合起来伤害到我，是因为，如果两者都以不同方式采取行动的话，我本不会死亡。

尽管（C7）此处给出的答案是对的，但是这个事例看起来提供了对（C6）的一个诘难。宣称在杀死我这一方面，Y并非在伤害我，这看起来荒谬。但是考虑一下

事例三。如前所述，X哄骗我喝下那种几分钟内便造成痛苦死亡的毒药。Y知道，如果他以一种其不可避免的副作用是我的即刻无痛苦死亡的方式行动，他能够救你的命。因为Y也知道我将痛苦地死去，所以Y以这种方式行动。

（C6）蕴含着Y应该以这种方式行动，因为他将不伤害我，而且他将带给我极大的利益。这结论是对的。既然Y的行动对我并非更糟，Y杀死我这一点在道德上是不相关的。X并未杀死我这一点在道德上也是不相关的。（C6）正确地蕴含着，X的行动是错的。尽管按照“杀”的通常用法X并未杀死我，但此处他是真正的凶手。X不但伤害到我，而且行动错误，因为这样一点是真的，就是如果他没有给我下毒的话，Y本来不会杀我。如果X以不同方式行动的话，我本不会死。Y并不伤害我，因为如果Y以不同方式行动的话，这本来也不会对我是否死亡带来什么不同。既然Y并不伤害我，而且极为有益于我，Y就是在做他应该做的。

正如这些评述所表明的，事例二并未提供对（C6）的什么诘难。在事例三中，（C6）正确地蕴含着Y应该像他所做的那样行动，因为他并不伤害我。在事例二中，Y的行动以相同的方式影响到我。所以我说在事例二中Y并不伤害我这一点是对的。在事例二中Y行动错误，是因为他有意识地成为合起来伤害到我的一个群体中的一员。

可能有人提出诘难说，如果这就是在事例二中Y行动错误的原因，Y在事例三中必定是在错误地行动。此处也可以认为Y有意识地成为合起来伤害到我的一个群体中的一员。

这个诘难表明需要另一个主张。在事例三中这一点是真的，那就是，如果X和Y两者都有所不同地行动的话，我本来不会受到伤害。但是这并不表明X和Y合起来伤害了我。这一点也是真的，即如果X和Y，以及弗雷德·阿斯泰尔[3]全都有所不同地行动的话，我本来不会受到伤害。但是这并不使弗雷德·阿斯泰尔成为合起来伤害到我的一个群体中的一员。我们应当主张

（C8）当某个群体合起来伤害到他人或者有益于他人的时候，这个群体是这样一个最小的群体，就其而言为真的是，如果他们全都有所不同地行动的话，他人本来不会受到伤害或者得到益处。

在事例三中，这个“群体”由X组成。就X而言这样一点是真的，就是如果他有所不同地行动的话，Y本来也会这样做，而且我本来也不会受到伤害。就Y而言则没有这样的主张为真。在事例二中，无论就X而言还是就Y而言这样的一点都不为真，就是如果他有所不同地行动的话，我本来不会受到伤害。只有当两者都不同地行动的时候，我本来才不会受到伤害。如果X和Y以及弗雷德·阿斯泰尔不同地行动的话，我本来也不会受到伤害。但是（C8）确当地蕴含着，弗雷德·阿斯泰尔不是合起来伤害到我的那个群体中的一员。这个群体由X和Y构成。

下面考虑一下

第三救援队：如前所述，如果4个人站到一个平台上，这将救起100个矿工的性命。有5个人站到这个平台上。

对于这5个人中的每一个人而言，鉴于其他人所做的，他的行动并不带来什么差异这一点是真的。如果他并未站到这个平台上，其他4个人也会救出100个矿工。尽管凭借他本身没有人会带来什么变化，但是这5个人合起来救了100个矿工。这个事例表明需要给（C8）添加某个进一步的主张。在这个事例中不存在合起来救出100个人的一个最小的群体。我将在第30节中回到这样一些事例。

还有我们应当考虑到行动集合的后果的第二种事例。这是一些协同难题。例子之一如下所示。

		你 做（1）	你 做（2）
我	做（1）	第二好	糟
我	做（2）	糟	最好

假设我们只把后果论应用于单一行动。那么我们将主张说，如果每一个已经做了对他来说可能的那些行动中的那个其后果是最好结果的行动，他就已经成功地遵循了C。正如我们先前所看到的，那么在

协同难题方面C将是不确定的。在这个事例中，既在倘若我们两者都做（2）时，又在倘若我们两者都做（1）时，我们才成功地遵循了C。

假设我们两者都做（1）。鉴于你所做的，我已经做了其后果最好的行动。如果我做（2）的话，结果就会更糟。相同的主张适用于你。如果我们两者都做（1），两者都成功地遵循C，但是我们没有产生出最好的可能结果。

后果论者应当主张

（C9）假设，鉴于他人所做的，我们各自已经尽其所能地使结果为好。那么各自就已经做出了正确的行动。但是我们合起来则可能已经行动错误了。如果我们合起来本来会使结果更好一些的话，情形就将是如此。

这是有关客观上的正确性或者有关什么事实上将使结果最好的一个主张。如果C包括这个主张，那么它在这类事例中将不再是不确定的。（当我们在决定去做什么的时候，我们应当代之以问什么是主观上正确的，或者鉴于我们的信念，什么将很有可能使结果最好。在绝大多数此类协同难题中，对我们每一个人而言，目标对准最好的结果主观上是正确的，因为这是其他人很有可能去做的。）[4]

27.忽视微小的几率

现在回到其中牵涉到许多人的那些囚犯的两难困境。时常有这样的主张，就是在这些事例中我们不能够诉诸我们行动的后果。这至少涉及另两个错误。

其中一个错误涉及的事例是，每一个利他主义的行动产生某种极大利益的几率极小。有时有这样的主张，就是在某个临界值之下极小的几率没有什么理性的或者道德的意义。

在有关选举的讨论中人们时常犯这个错误。尽管选举不是一个纯粹的囚犯的两难困境，但是它可以说明这个错误。许多作者声称，在

全国性的选举中，一个人仅凭诉诸自己行动的后果不能证明自己所投一票的正当性。但这经常是虚假的。假设，如果我参加投票，这将涉及某些代价，而且除了对赢得选举的人的可能具有影响之外将不带来什么利益。按照这些假设，我所投的一票以利己的标准衡量并不能得到正当性的证明。但是它以后果论的标准衡量却时常得到正当性的证明。当我不能够预测我的行动的后果的时候，C告诉我要去做不管怎样会产生最大的预期的利益的事。我的行动的预期利益就是可能的利益与我的行动产生这个利益的几率相乘所得。我可能能够诉诸这个利益来证明我所投的一票的正当性。[5]

考虑一下美国的总统选举。如果我参加投票，我所投的选票带来不同结果的几率非常小。按照一种估算，如果我在那些选情不定且会以微小差距决定选情的大州之一进行投票，我将带来不同结果的几率会是大约一亿分之一。（这个估算相当困难。而且不应当假定任何一个投票模式都会像别的一样。但是有几位作者同意，这个几率大约是亿分之一。[6])

姑且称两位总统候选人为优先生和劣先生。而且假设如果下届当选总统是优先生的话，美国人民总体上将获益。当然有一些美国人将有一些损失。对这些美国人而言劣先生胜出本会更好一些。但是这些美国人——富裕的少数人——所蒙受的损失将被其他所有美国人的获益所超过。这也就是优先生之所以是更好的候选人的原因。如果他胜出的话，将产生减去负担后的更大的净效益总额。美国人的平均净利益是这个总额被美国的人口所除得到的数目。为简单起见，我忽略了对非美国人的影响。如果我的票影响到结果的几率是亿分之一，我所投下的一票的预期利益如下式所示：

$$\left(\frac{\begin{matrix}\text{优先生胜出}\\\text{给美国人带来}\\\text{的平均净利益}\end{matrix}\times\begin{matrix}\text{美国人}\\\text{的人数}\end{matrix}}{1\text{ 亿}}\right)-\left(\begin{matrix}\text{我和他人为}\\\text{我的投票}\\\text{付出的代价}\end{matrix}\right)$$

既然美国有2亿人口，这个总额很可能是正的。如果优先生的胜出总体上会带给美国人一个大于为我投票所花的一半代价的净利益，就会如此。要怀疑这一点，我必须是相当愤世嫉俗的。同样的评述适用于许多别的公共的善，不仅适用于后果论者，也适用于利他主义者。如果一个利他主义者并不忽视非常微小的几率，他将时常具有据以作出贡献的道德理由。他给予他人的所预期的利益会大于他的贡献所付出的代价。

可能会有诘难说，考虑非常微小的几率是非理性的。当我们的行动所影响到的不过是几个人的时候可能如此。但这是因为此处的风险相对而言是低的。考虑一下造成意外死亡的风险。给予任何杀死一个人的想法以百万分之一的几率可能是非理性的。但是，如果我是一位核工程师，给予杀死一百万人的想法以同样的几率是非理性的吗？这不是我们绝大多数人所认为的。我们恰当地认为，这样的一些机会应当予以考虑。假设核工程师们的确忽视了处于百万分之一极限值或低于这个极限值的所有那些几率。那么情形可能是这样的，就核反应堆的许多组件中的每一个组件而言，任何一天都有百万分之一的机会出现故障，而造成一场大灾难。对那些设计反应堆的人们而言忽视这样的微小几率显然是错的。如果有许多核反应堆，而每一个都有这样的一些组件，那么不用多久这种百万分之一的风险就会经历百万次。核灾难很快就会出现。

当风险相当高的时候，不应当忽视任何几率，无论这个几率是多么的小。当一个几率将有许多次会发生的话也是如此。在这两类情形中，每一个微小的几率都应当予以应有的对待，并且考虑进预期利益的计算之中。我们通常会忽略一种非常细小的几率。但是当我们可能影响到数目非常大的人群的时候，或者当这个几率出现的次数将会非常大的时候，我们就不应当这样做。这些庞大的数目大体上抵消了这个几率的微小性。

如果一个行动很可能或者肯定带给他人非常小的利益，类似的情况同样适用。当这些利益涉及的人数相当庞大的时候，我们不应当忽视它们。这个庞大的数目大致抵消掉了这些利益的微小性。从而利益的总和可能是巨大的。

这两点并非同等可信的。非常细小的利益可能是觉察不到的。而且主张“觉察不到的利益”不是一种利益这一点是可信的。但是主张非常细小的几率不是一种几率却并不令人信服。

28.忽视细小的或者觉察不到的后果

道德数学中的第三个错误是忽视要么会影响到许多人，要么会出现许多次的那些非常小的几率。第四个和第五个错误是忽视对非常大的数目的人所产生的非常细小的和觉察不到的那些后果。这是些类似的错误，可以用同样的论证加以批判。但是觉察不到的后果产生一个额外的问题。

我无需对两个错误都加以陈述。除了“非常细小的”取代了“觉察不到的”之外，第四个错误与第五个一模一样。一些人认为：

> （第五个错误）如果某个行动对他人产生觉察不到的后果，这个行动不能因为具有这些后果而在道德上是错误的。一个行动不能够因为其对他人的后果而在道德上是错误的，如果这些人中没有什么人会注意到任何差异的话。同样，如果某个行动对他人会产生觉察不到的后果，这些后果不能够使该行动成为某个人所应该做的。

有一类觉察不到的后果是没有争议的。我可能以一种察觉不到的方式对你造成严重伤害。我所给予你的辐射剂量可能是导致许多年后夺去你生命的那种癌症的不为人知的原因。尽管该原因不为人所知，此处的后果却是可以觉察到的。但是在我将予以考虑的那些事例中，后果则是觉察不到的。

首先考虑一下格洛弗所表述的那种事例的一个变体。[7]

水滴。一大批伤员躺卧在露天沙漠中，忍受着难耐的干渴。我们则是相等数量的一批利他主义者，每一个人都有一品脱水。我们可以把这些水注入一个运水车中。运水车将被开到沙漠之中，我们的水将在这些伤员中平均地分配。通过添加自己的那一品脱水，我们每一个人会使得每一个伤员多喝一点点水——或许只有额外的一滴水。即便对一个非常干渴的人而言，这些额外水滴中的每一滴水也不过是一个非常细小的利益。对每一个人的后果可能是觉察不到的。

设若每个人所得到的益处只是解除他们痛苦难当的干渴。这些人的健康不会受到什么影响。既然那些益处只不过是解除痛苦，那么它们就是这样一种益处，就是说，对其可以极为令人信服地主张说，它们要成为益处则必须是可以觉察到的。

首先假设，因为数目不是非常庞大，我们每一个人给予每一个伤员的利益尽管非常细小，却是可以觉察到的。如果我们犯第四个错误，我们就会认为这样细微的利益不具有什么道德上的意义。我们认为，如果某个行动会带给他人这样的一些细微的利益，这不能够使这个行动成为某个人所应该做的。我们被迫得出结论，我们大家都不应该添加他的那一品脱水。这显然是错误的结论。

其次假设，有一千名伤员和一千名利他主义者。如果我们把我们的那些品脱的水注入那辆运水车，我们每一个人将使每一个伤员多喝到千分之一品脱的水。这些人或许会注意到喝不到什么水与喝到千分之一品脱的水之间的差异。从而允许我们追问这样一个问题，“如果这些人至少喝到十分之一品脱的水，他们会注意到多喝到千分之一品脱的水的后果吗？”我将假定答案是否定的。（如果答案是肯定的，我们只需要假设有更多的伤员与利他主义者。必定有一品脱的某个分数的后果小到无法觉察。）

假设一百位利他主义者已经把他们的水注入那辆运水车中。每一个伤员至少喝到十分之一品脱的水。我们是另外九百位利他主义者，每一个会添加自己的那一品脱水。接着假设我们犯第五个错误。我们相信，如果某个行动对他人具有觉察不到的后果，这些后果不能够使

这个行动成为某个人所应该做的。如果我们相信这一点，我们就不能够解释我们每一个人为什么应该添加他的那一品脱水。

可能会有这样的说法："如果我们重新表述添加每一品脱水的后果，我们能够避免这个难题。我们无需主张这给予每一个伤员千分之一品脱的水。我们本可以主张这给予每个伤员一品脱的水。"

这个主张是错谬的。水将在所有这些伤员之间进行均等的分配。当我添加我的那一品脱水的时候，后果是额外的一个人得到一整品脱的水吗？如果我没有添加我的那一品脱的水的话，会有某个人不是得到一整品脱的水而是什么也没有得到吗？这些没有哪一个是真的。对我的行动的后果只有一种正确的表述。它给予一千位伤员中的每一个额外的千分之一品脱的水。

下面可能有人说我们应当诉诸总体份额的观点。按照这个观点，每一个人所贡献的份额相当于一个人从一品脱水中所得到的利益。但是我们不能够诉诸这个观点，因为我们在第25节中看到它可能蕴含一些荒谬的结论。

我们所能够诉诸的是有关我们合起来做什么的主张。我们能够主张

> （C10）当（1）最好的结果会是其中人们得益最大的那个结果的时候，且（2）某个群体中的每一个成员会以某种方式行动的时候，且（3）如果他们有足够多的人以这种方式行动，他们会给这些其他人带来益处的时候，且（4）如果他们全都以这种方式行动的话，他们会给这些人带来最大的益处的时候，且（5）他们每一个人既知道这些事实又认为他们足够多的人将以这种方式行动的时候，那么（6）他们每一个人应该以这种方式行动。

我们每一个人会给予那一千位伤员额外的千分之一品脱的水。如果我们有足够多的人以这种方式行动，这将给这些人中的每一个人带来益处。而且，如果我们全都以这种方式行动我们给这些人带来的益处最大。我们知道这些事实，而且我们知道我们足够的人——一百个人——已经以这种方式行动。（C10）正确地蕴含着，我们每一个人应该以这种方式行动。

现在回忆一下第五个错误。按照这个观点，如果这些后果是觉察不到的，一个行动就不能够因为其对他人的后果而成为对的或者是错的。刚刚表述的那个事例驳倒了这个观点。显然，在这个事例中我们每一个人应当把自己的那品脱水注入那辆运水车中。我们每一个人应当使每一个伤员多喝到千分之一品脱的水。我们每一个人应该以这种方式影响到每一个伤员，即使这些后果是觉察不到的。我们可能认为，因为这些后果是觉察不到的，我们每一个人没有带给什么人益处。但是，即便每一个人没有带给什么人益处，我们合起来将极大地有益于这些伤员。所有我们大家的行动的后果是觉察得到的。但是，我们大大地缓解了这些人们的极度干渴。

后果论者可能诉诸数个原则。他们可能从而认为，在某些事例中最好的结果不是人们在其中受益最大的结果。要涵盖这样一些事例，他们可以主张

> （C11）当（1）如果某个群体中有足够多的人以某种方式行动，则某个群体中的那些成员会使结果更好一些的时候，且（2）如果他们全都以这个方式行动他们会使结果最好的时候，且（3）他们每一个人既知道这些事实又认为他们有足够多的人将以这种方式行动的时候，那么（4）他们每一个人应该以这种方式行动。

非后果论者认为，在一些事例中我们应当力求产生最好的结果。在这样一些事例中，他们可以诉诸（C11）。如前所述，在某些事例中（C11）本身并不给出正确的答案。我们会不得不添加一个更加复杂化的主张。我将略去这些复杂之处。[8]

正如我在第26节中所表明的，有两类事例我们需要诉诸后果，不仅是单一行动的后果，而且是行动集合的后果。当（1）我们行动的后果是多因素决定的时候或者（2）当我们面对协同难题的时候，我们需要依仗这样的诉诸。我们现在考虑的事例是（3）每一个人的行动将对他人具有觉察不到的后果。这可能是我们在其中需要诉诸行动

集合的后果的第三类事例。我们是否需要做出这样的诉诸部分取决于对另一个问题的回答。

29.可能存在觉察不到的伤害和受益吗？

可能遇到这样的诘难："你主张说，一千个利他主义者中的每一个都应当注入他的那一品脱水，因为这是那些伤员得到最大利益的方式。这个主张是错谬的。假设那些利他主义者中有一个并未注入他的那一品脱水。那些伤员得益少了吗？他们没有。他们所喝到的水稍微少一些。但是这个后果是觉察不到的。既然该后果是觉察不到的，这些人的利益就不能算是更少了一些。"

这个诘难假定不会有觉察不到的利益。如果我们作出这个假定，我们则面对更广泛的一个难题中的部分，这个难题有复合三段论难题（Sorites Problem）、王浩悖论（Wang's Paradox）和谷堆悖论（Paradox of the Heap）等不同的称呼。

在我们所谈论的事例中，利益在于解除令人难耐的干渴。如果每一个人得到一品脱的水，他的干渴所造成的痛苦就变得小一些。他的痛苦的程度就将小一些。我们的难题如下。我们假定

（A）某个人的痛苦不可能觉察不到地变得更好或者更糟。如果这个人可能不会注意到任何差异，那么其痛苦就不能够变得要么不那么糟糕要么更糟。

而且可以令人信服地假定

（B）至少像……一样糟糕和不比……更糟糕当被应用于描述痛苦的时候是传递性的关系。于是，如果某个人的痛苦在结果（2）中至少像结果（1）中一样糟糕，而且他在结果（3）中的痛苦至少像结果（2）中一样糟糕，那么他在结果（3）中的痛苦至少像结果（1）中的一样糟糕。

一百个利他主义者已经把他们那些品脱的水注入了运水车。每一个伤员至少将会喝到十分之一品脱的水。他们不会注意到任何额外的千分之一品脱的水的后果。在不同的可能结果中，不同数目的利他主义者后来将把他们那些品脱的水注入运水车。且让我们通过引用作出贡献的人们的数目来指涉这些结果。于是，如果没有什么别人作出贡献，这将产生结果100。

假设又多一个利他主义者作出贡献。每一个伤员喝到的水将多一些，但是多出来的量是如此之少，以至于他并不能够注意到这一点。根据（A），每一个人的干渴的痛苦不可能变得更轻一些。在结果101中，每一个人的痛苦必定至少会像结果100中一样糟糕。接着假设第二位利他主义者添加上他自己的那一品脱水。像前面的情况一样，那些人中没有什么人能够注意到这个差别。根据（A），每一个人干渴的痛苦不可能变得更轻一些。在结果102中每一个人的痛苦必定至少像结果101中一样糟糕。根据（B），每一个人在结果102中的痛苦必定至少像在结果100中一样糟糕。如果又有第三位利他主义者作出贡献，相同的主张同样适用。在结果103中，每一个人的痛苦必定至少像在结果100中一样糟糕。这些主张适用于每一额外的利他主义者作出贡献的情形。假设我们都作出贡献。所出现的结果便是结果1000，其中每一个伤员喝到一品脱的水。（A）和（B）合起来蕴含着，每一个人的痛苦必定至少像在结果100中一样糟糕。喝到一品脱水，而不只是十分之一品脱的水，在解除每一个人干渴的痛苦方面并不能够起任何作用。既然这个结论是荒谬的，我们必须要么拒斥（A）要么拒斥（B）。

应当予以拒斥的是哪一个？我拒斥的是（A）。我认为，凭小得无法注意到的量，某个人的痛苦能够变得轻一些或者不那么糟糕一些。如果某个人更多留心其痛苦，或者有更强烈的去除这个痛苦的愿望，那么这个人的痛苦在具有道德相关性的意义上更糟。我认为，即便某个人不能够注意到任何差异之处，但是某个人能够对他的痛苦留

意稍少一点，或者对去除这个痛苦的愿望稍弱一点。更加宽泛一些说，能够有觉察不到的伤害和觉察不到的受益。在许多别的种类的事例中，当人们报告他们的经验的本质的时候，业已表明犯有非常细小的错误。我们为什么应当假定在去除某种痛苦的渴望强度方面他们不会犯这样的错误?

假设你拒斥这些主张，而且继续接受（A）。那么你必须拒斥（B）。要避免上述所达到的那个荒谬结论，你必须承认，当应用于疼痛的时候，“至少像……一样糟糕”和“不比……更糟”不是传递性的关系。而且拒斥（B）具有像拒斥（A）那样的一些蕴含。现在你必须承认，你的行动可能因为其对别人的疼痛所产生的后果而是错误的，即便你的行动一点没使这个人的疼痛更糟。它们合起来可能具有这个后果。每一个行动可能是错误的，即便其后果是觉察不到的，因为它是合起来造成这个人的疼痛糟糕得多的一个行动集合中的一个。

考虑一下

过去的糟糕日子。一千个施虐者有一千个受害者。在每一天之始，每一个受害者都已经感到轻微的疼痛。每一个施虐者把某种仪器上的开关按一千次。开关的每按一次都以一种觉察不到的方式影响到某个受害者的疼痛。但是在每一个施虐者按他的开关一千次之后，他就给他的受害者施加了剧烈的疼痛。

假设你犯了第五个错误。你认为，如果一个行动的后果是觉察不到的，就不能够因为其对他人的后果而是错误的。那么你得出结论说，在这个事例中没有哪一次按开关是错误的。这些施虐者中没有哪一个行动错误。这个结论是荒谬的。

那些施虐者何以是在错误地行动?一种解释是诉诸每一个施虐者所做的事情的总体后果。每按开关一千次，这些行动合在一起对他的受害者施加了剧烈的疼痛。

下面考虑一下

无害的施虐者。在过去的糟糕日子中，每一个施虐者对一个受害者施加了剧烈的痛苦。现在事情出现了变化。一千个施虐者各按一下按钮，从而开启一千件仪器上的开关各一次。受害者们遭受同样剧烈的痛苦。但是施虐者中没有哪一个使任何受害者的痛苦可觉察地更糟一些。

我们在此能够诉诸每一个施虐者所做的事情的总体后果吗？这部分取决于我们是否拒斥（A），认为某个人的痛苦能够变得比觉察不到地更糟。如果我们这样认为，我们可以主张："通过按下按钮，每一个施虐者造成每一个受害者遭受略微多一点的痛苦。在每一个人身上的后果是轻微的。但是，既然每一个施虐者增加了一千个受害者所遭受的痛苦，那么他就对他们施加了巨大的受苦总量。既然那些受害者所遭受的痛苦像他们在过去的糟糕日子中所遭受的一样多，这些施虐者所犯的错误就像他们在过去的糟糕日子中所犯的一样。在过去的糟糕日子中，每一个施虐者对一个受害者施加了巨大的痛苦总量。每一个无害的施虐者对这一千个受害者施加了同等大的受苦总量。"

相反假设我们接受（A），认为痛苦不可能觉察不到地变得更糟。那么我们必须承认，每一个无害的施虐者没有使哪一个人的痛苦更多一些。我们不能够诉诸每一个施虐者所做的事情的总体后果。按照我们的观点，没有哪一个施虐者伤害到任何一个人。

即使他们中没有哪一个伤害到任何一个人，那些施虐者显然也是在错误地行动。如果我们不能够诉诸每一个施虐者所做的事情的总体后果，那么我们必须诉诸那些施虐者合起来所做的后果。即便他们没有哪一个造成任何痛苦，他们合起来对一千个受害者施加了巨大的痛苦。我们可以主张

（C12）当（1）如果人们受痛苦更多则结果会更糟的时候，且（2）某个群体中每一个成员以某个特定方式行动，且（3）倘若他们有足够多的人以这种方式行动的话他们会造成他人痛苦，且（4）倘若他们全都以这个方式行动他们会造成这些人痛苦最大，且（5）他们每一个人既知道这些事实又相信他们中足够多的人将以这个方式行动，那么（6）每一个人倘若以这个方式行动就会是在错误地行动。

某人可能再次诘难说：“在无害的施虐者那种事例中，（4）并不为真。如果这些施虐者全都每人按一次开关的话，他们并不造成他们的受害者痛苦最大。假设他们中有一个人没有按开关。那些受害者中没有哪一个会注意到任何差别。既然一个疼痛不能变得觉察不到地更少一些，如果那些施虐者中有一个并未去做的话，那些受害者不会痛苦更少一些。”

正如我所说到的，这个诘难引发了那个著名的复合三段论难题。如果我们接受（A），我们对这个诘难的回答必定牵涉到对这个难题的解决方法。由于这个难题既难以解决又引发的是些与伦理学没有关系的问题，所以我在此将不予讨论。[9]

如果我们接受（A），我们对无害的施虐者的诘难就必须既是复杂的又是解决复合三段论难题的。如果我们拒斥（A），我们的诘难就会是简单的。我们会主张说，每一个施虐者对那些受害者施加了巨大的痛苦总量。

就这两个解释而言，哪一个更好一些？即使我们拒斥（A），我们给出简单一些的解释可能是错误的。是否如此取决于对另一个问题的回答。考虑一下

> 单独的施虐者。一天早晨，施虐者中只有一个出场工作。碰巧为真的是，因为自然原因，每一个受害者已经遭受相当剧烈的痛苦。这个痛苦的程度相当于那些开关被按500次那样糟糕。由于知道这个事实，那个单独施虐者在所有的机器上按了一次。其后果就像所有的施虐者都行动的那些日子的情况一样。更精确些说，后果就像每一个开关被按第501次一样。那个单独施虐者知道后果正是这样。他知道他不是在使任何一个受害者的疼痛可以觉察得到地更糟。而且他知道他不是合起来行动的一个群体中的一员。

该单独施虐者是在错误地行动吗？假设我们认为他不是。那么我们不能够诉诸所有的施虐者都行动的那种事例下的那个简单一些的诘难。我们不能够主张说，每一个人都是在错误地行动，因为他是在给他人施加巨大的痛苦总量。如果这就是为什么每一个人是在错误地行动的原因，那么，该单独施虐者必定是在错误地行动。他以相同的方

式行动，而且具有同样的后果。如果我们认为该单独施虐者并不是在错误地行动，那么，我们必须在所有的施虐者都行动的事例中给出别的诘难。我们必须主张，每一个人都是在错误地行动，因为他是合起来给他们的受害者施加巨大痛苦的一个群体的成员。

我认为该单独施虐者是在错误地行动。不管他是与其他行动者联合还是与大自然一起，产生糟糕的后果怎么能够造成道德上的差异？[10]我因而在两种事例中都倾向于诉诸单独行动的后果。

一些人并不同意。即使我们认为能够有觉察不到的伤害和受益，但诉诸群体合起来所做的事情可能更好一些。这一诉诸争议性更少一些。

（如果单独的施虐者并不是在错误地行动，宣称我们有些人在道德数学方面犯下五大错误可能并不公平。按照这个观点，第五个错误只是第二个错误的一个特例。但这很少会造成一种实际上的差异。）

在本节中我追问了是否可能存在觉察不到的伤害和受益。我倾向于给出肯定的回答。如果我们给出否定的回答，我们必须舍弃这样一个主张，就是，当应用于伤害和受益时，"至少像……一样糟糕"和"不比……更糟"是传递性的关系。我还表明了，无论我们接受哪一个答案都没有多大差异。按照两个答案中的无论哪一个，我们必须舍弃我们所称的第五个错误。我们必须放弃这样一个观点：如果一个行动的后果是觉察不到的，就不能够因为其对他人所产生的后果而要么是对的要么是错的。

30.多因素决定

现在回到有关那些品脱的水和伤员的事例。且让我们对这事例再添加一些特征。假设，你在水车开到这些伤员那里之前带着另一品脱的水赶到。伤员们所需要的水超过单单的一品脱。在喝下这品脱水之后他们充满剧烈痛苦的干渴没有得到完全解除。但是水车只能够装下

一千品脱的水。现在水车已满。如果你添加上你的那一品脱的水，这只会造成一品脱的水溢出来。

你没有什么道德理由来添加你的那一品脱的水，因为这只会造成一品脱的水被浪费掉。根据（C10），如果这会使你成为合起来给他人带来利益的一个群体中的成员，你就应该添加上自己的那一品脱的水。我们可能认为，如果你添加上你的那一品脱的水，你并不就是合起来给那些伤员带来利益的一个群体中的成员。即使你的那一品脱的水中的一些，可能被这些伤员们喝到。而且即使你就是在像其他利他主义者所做的那样行事。但是我们也许会主张说，"与其他利他主义者们不同，你并未给予每一个伤员额外的千分之一品脱的水。你的行动对这些人所得到的水量并没有什么后果。"

事情并不如此简单。如果你添加上你的那一品脱的水，这将是一个涉及多因素决定的事例。就你而言这一点的确为真，就是如果你并未添加你的那一品脱水的话，这并不会对那些人所得到的水量带来什么差异。但是，既然你添加了，这一点就是真的，就其他每一个利他主义者而言情况同样如此：就是如果这些利他主义者中任何一个人没有添加的话，并不会对这些人所喝到的水量造成什么差异。在你到来的时候水车并没有满，而你的那一品脱的水会使水车满起来。就你而言为真的，就其他每一位利他主义者而言也为真。因此，你是合起来给这些伤员带来利益的群体中的一员这一点为真。

我们必须再次诉诸行动者所知道的或者他们有理由所相信的。假设，其他利他主义者没有理由相信你会带着你额外的那一品脱的水到来。每一个都应该把他那一品脱的水注入水车。这是因为每一个人有充分的理由相信他会是下述情形为真的一个群体的成员：（1）他们合起来给伤员们带来利益，而且（2）如果他们全都注入他们的那些品脱的水的话，他们给这些人带来的利益最大。当你到来的时候，你知道该水车已满。你没有什么理由添加，因为你知道你不会是这样的

一个群体中的一员。如果你添加水的话，你将成为过大的一个群体中的一员。我们应当主张

> （C13）假设有某个群体，他们通过某种方式采取行动一起给他人带来利益。如果某一个人认为，如果他加入的话，这个群体要么是过大，要么会过大，他没有什么道德上的理由加入这个群体。如果一个群体的一个或者多个成员没有行动不会减少该群体给他人带来的利益这一点为真，那么这个群体就是过大的。

如果你添加上你的那一品脱水的话，这将使这个利他主义者的群体成为过大的。如果你并不添加上你的那一品脱的水，这个群体将不是过大的。这是一个特别的划界事例，（C13）也涵盖更常见的那些已经过大的事例。

31.合乎理性的利他主义

道德数学中的第五个错误是这样一个信念，就是觉察不到的后果在道德上不可能是重要的。这是一个非常严重的错误。当所有的无害的施虐者行动的时候，即便他们每一个并没使什么人可以觉察到地变得更糟，他们仍旧都是在非常错误地行动，这一点为真。就我们而言相同的情况也会是真的。我们应当不再认为，如果一个行动并没使什么人可以觉察到变得更糟，该行动就不能够因为其对他人的后果而是错误的。我们每一个行动都可能因为对其他人的后果而是非常错误的，即便这些人中不会有什么人注意到这些后果。我们的行动可能合起来使这些人的处境极其糟糕。

第四个错误同等严重。如果我们认为微不足道的后果在道德上可以予以忽略，那么我们可能时常会使人们的处境极其糟糕。回忆一下那个渔民的两难困境。凡存在过度捕捞或者渔业资源总量下降之处，就每一个人而言如果力求捕获更多则能够更好一些，而如果全都这样做，那么对每一个人来说则更糟。考虑一下

渔民何以造成灾难。有许多渔民，他们靠在某个水面浩瀚的大湖上各自捕鱼维持生计。如果每一个渔民不限制自己的捕获量的话，他将在接下来为数不多的季节中捕获到更多的鱼。但是他将从而大大降低总体捕获量。由于有许多渔民，如果每一个人都不限制他的捕获量，他只将微不足道地影响到其他每一个人所捕获的数量。渔民们认为，这样微不足道的后果道德上是可以忽略的。因为他们这样认为，即便他们从不做他们认为是错误的事情，他们并不限制他们的捕获量。每一个因而增加了自己的捕获，但是却造成总体捕获量的更大幅的下降。因为他们都以这种方式行动，结果是一场灾难。在几个季节之后，捕获量全都少了许多。他们将不能维持自己的生计和抚养子女。

如果这些渔民知道这些事实、具有更多的利他主义而且避免第四个错误的话，他们本会避免这个灾难。每一个人知道，不管他人怎么做，如果他不限制自己的捕获量，这对他自己来说在一定程度上将更好一些。而且每一个人知道，如果他以这个方式行动，对其他每一个人的后果将是微不足道的。但是渔民们不应当认为，这些微不足道的后果在道德上是可以忽略的。他们应当认为，以这样的方式行动是错误的。

如同前面一样，我们有两个方式来解释为什么这些行动是错误的。我们会诉诸每一个人的行动的总后果。每一个渔民知道，如果他并不限制自己的捕获，他将捕获到更多的鱼，但是他将大为降低总捕获量。为了自己的小增益，他给他人施加了一个要大得多的总损失。我们会主张说，这样的行动是错误的。这个主张并不假定能够有觉察不到的伤害和利益。因而，较之有关每一个无害的施虐者的所作所为的那个相应的主张，它的争议性更少一些。

我们的另一个选择是诉诸这些渔民们合起来做了什么。每一个渔民知道，如果他和其他所有的人不限制他们的捕获量，他们合起来将给他们自己造成巨大的总体损失。理性的利他主义者会相信，这些行动是错误的。他们会避免这个灾难。

可能有人说：“理性的利己主义者也会如此。每一个人知道，如果他并不限制他的捕获量，他就是在他们自己身上施加巨大损失的一个群体中的一员。即便按照利己的标准衡量，以这种方式行动也是非理

性的。”正如我将在下一章中所论述的，这个主张是得不到正当性证明的。每一个人知道，如果他不限制他的捕获量，对他自己而言这将更好一些。无论他人做什么，都是如此。当某个人做他所知道的对自己更好一些的事情的时候，不能够这样说，他的行动以利己标准衡量是非理性的。

下面回忆一下：

> 上班族的两难困境。假设我们生活在一座大都市的郊区。我们能够开车或者乘公共汽车来回上下班。由于没有公共汽车车道，超额的交通量就像阻塞小汽车一样阻塞公共汽车。我们因而会知道下述情形为真。当我们绝大多数人开车上下班的时候，如果我们中的任何一个人开车而不是乘公共汽车，他将从而为自己省下一些时间，但是他将给他人施加大得多的总时间损失。这个后果会得到扩散。每一个人可能造成一百个他人耽误20秒，或者造成一千个他人耽误2秒。我们绝大多数人会把这样的一些后果当作是微不足道的，以至于在道德上是能够予以忽略的。我们然后会认为，在这个上班族的两难问题中，即使一个理性的利他主义者也能够具有正当性地选择开车而不是乘公共汽车。但是如果我们绝大多数人都作出这一选择，那么我们每天全都将被耽误很长时间。

理性的利他主义者会避免这个结果。如同前面，他们会诉诸每一个人的所作所为的后果，或者诉诸所有的人合起来所作所为的后果。每一个人为自己省下一些时间，却是以给别人施加大得多的总时间损失为代价。我们会主张说，以这种方式行动是错误的，即便对其他每一个人的后果会是微不足道的。我们会代之主张说，这个行动之所以是错误的，是因为那些以这种方式行动的人们合起来对每一个人造成巨大的时间损失。如果我们接受这些主张中的任何一个，而且具有充分的利他主义，我们就会解决上班族的两难困境，每天为我们自己省下大量时间。

同样的推理适用于其他无数事例。再考虑一个汽车废气净化设备的例子。我们会认为，如果自己省下维修这个设备的花费，其结果是对某个单独的人造成巨大的空气污染，那么这是错误的。但是，如果其结果只是微不足道地或者觉察不到地增加众多人中的每一个人所遭受的空气污染，我们许多人则不会认为这是错误的。这是在许多大都

市中的实际情况。如果我们中没有人造成这样的污染，对我们大家而言或许要好得多。但是，要相信我们是在错误地行动，我们中的许多人尚需要改变观念。我们必须不再相信下述这一点：如果一个行动的影响是微不足道的，或者是觉察不到的，那么该行动之错则不可能在于其对别人产生的影响。

随着条件的变化，我们可能需要在道德思考方式方面作某些改变。我一直在为这样的一种变化作论证。常识道德在小共同体中发挥的作用最好。当我们的人数不多的时候，如果我们给予他人巨大的总体利益或者强加给他人巨大的总体伤害，我们必定是在以一些重要的方式影响到他人，这些要么会是我们感激的根据，要么会是愤慨的根据。在小共同体中，如果没有什么人对我们所做的事带有明显的抱怨之情，或者对我们所做的事没有什么愤慨的根据，那么我们不可能伤害到他人这一主张就是可信的。

直到20世纪，人类中的绝大多数都生活在小共同体中。每一个人所做的只会影响到一些人。但是现在情况已经发生了变化。我们每一个人现在都能够以无数的方式影响到无数的他人。我们能够对数以千计或者数以百万计的人产生尽管微小却是真实的影响。当这些后果被广泛地扩散的时候，它们可能是微不足道的或者是觉察不到的。现在这一点有了很大的差异：我们是否继续相信，除非有人具有明显的感激或者愤慨的根据，否则我们就不可能极大地给他人带来伤害或者收益。

当我们继续相信这一点的时候，即便我们在乎对他人的后果，我们也可能无法解决许多严重的囚犯的两难困境问题。为我们自己的或者家庭的一些小利益之故，我们每一个人可能剥夺他人大得多的总体利益，或者给他人带来大得多的总体损失。我们可能认为这是允许的，因为这对那些他人中的每一个的后果将是微不足道的或者是觉察不到的。如果这就是我们所认为的，那么我们的所作所为对我们大家而言将经常是糟糕得多的。

如果我们充分在乎对他人的影响，而且改变我们的道德观点，我们就会解决这样一些难题。只是问"我的行动将伤害到他人吗？"这个问题是不够的。即便答案是否定的，我的行动仍然可能因为其后果而是错误的。当单独考虑的时候它所具有的后果可能并非是其惟一相关的后果。我应当追问，"我的行动是合起来将给他人造成伤害的行动集合中的一个吗？"答案可能是肯定的。对他人的伤害可能是极大的。如果是这样，我可能是在非常错误地行动，就像那些无害的施虐者一样。如果我们对他人的关注是要产生化解我们所面对的绝大多数囚犯两难困境问题的解决方法，即产生我们面对的绝大多数事例的解决方法——如果我们每一个人而不是没人做将会对自己、对自己的家庭或者对自己所爱的人们更好的事情，这对每一个人而言将会更糟，而且时常是糟糕得多——那么，我们就必须接受这个观点。

---

[1]参见，例如，Ewing，Miller and Sartorius，Meehl，或者非常有影响的Olson。第64页和第159页。

[2]我把这些要点归功于马丁·霍利斯所作的一个尖锐的评论。

[3]弗雷德·阿斯泰尔（Fred Astaire，1899—1987），美国舞蹈家和电影演员。以舞姿优雅出名。——译者

[4]（C9）不同于里根所提出的那个更加复杂的合作后果论。（C9）不是修正行动后果论者有关我们什么时候各自正确行动的主张，而只是添加一个关于我们何时正确地行动的主张。（这些主张可能看起来不一致。如果我们各自已经正确地行动，何以我们一道却已经错误地行动？但此处没有什么矛盾。对于那些后果论者，如果一个行动在境遇中具有最好的可能后果，那么它就是客观上正确的。当我们问各人是否行动正确的时候，境遇包括他人所做了的事。当我们问我们是否正确地行动了的时候，则并非如此。）

[5]参见Meehl。

[6]参见Meehl，Riker and Ordeshook，以及Mackie（3）。

[7]Glover（2），第174—175页。Glover写道：

可能有人认为，在绝对临界值和识别临界值之间……没有什么差异。一些人试图同化那个缺电情形与那个投票情形。在缺电情形中，当我所做的伤害分布到那个共同体的时候是低于那个识别临界值的。那些把这两类临界值一视同仁的后果论者们得出结论说，除了副作用之外，我是否用电并不重要。这种提法是说，所造成的伤害算作零。

但是针对这一点，我想要论证的是，在这样一些情形中所造成的伤害应当算作一个可识别单位的分数，而不是算作零。让我们称之为可分性原则（principle of divisibility）。它说的是，在伤害是个程度问题的那些情形中，亚临界值行动在它们造成伤害的那个程度上说是错误

的，而且在必须有像我的行动那样的100个行动来造成一种可觉察的伤害的情形中，我已经造成了可觉察的那个伤害的百分之一。
假设一个村庄有100个正在吃午饭的部落成员。100个饥肠辘辘的武装匪徒突袭了这个村庄，匪徒用枪相逼，各自夺取一个部落成员的午饭并且吃掉。这些匪徒然后逃逸，各自对一个单一的部落成员做出了一种可以觉察的伤害。接下来的一个星期，这群匪徒试图故伎重演，但是受到对这样袭击的道德性的一些新发现的怀疑的困扰。他们的怀疑最终被他们那群人中一个并不相信可分性原则的人所打消。他们然后袭击了那个村庄，捆绑起那些部落成员，并且看着他们的午餐。正如预期的那样，各个碗中有100粒烤熟的豆子。源于一个烤熟的豆子的快乐低于那个识别临界值。各个匪徒不是像上个星期那样一下吃掉一整碗食物，而是各自从每个碗中取走一粒豆子。他们在吃掉所有的豆子之后离开，所幸没有造成什么伤害，因为各自对各个部落成员所做的并不多于亚临界值伤害。那些拒斥这个可分性原则的人们不得不同意。
这一章，特别是我后来的无害的施虐者们，完全脱胎于这个出色的例子。

[8]再次参见Regan。

[9]（此注为1987年所加）在他刊载于1986年7月《伦理学》之中的那篇文章里，B·格鲁察尔斯基表明我对这个问题的最初讨论是混乱的。也请参见Dummett，Peacocke（1），以及Forbes（2）。

[10]假设送给那些伤员的水不是来自那些别的行动者而是来自雨水。这会撤除我添加上我的那1品脱水的理由吗?

第四章

直接自败的理论

我们常常面对多个囚犯的两难困境。以下一点通常是真的，那就是，如果我们每一个人而不是没人做将对自己、对自己的家庭或者对自己所爱的人更好的事情，这对我们大家而言将更糟。如果我们每一个人倾向于以这种方式行动，这些情形就产生出一个实际上的难题。除非出现某种变化，否则结果对我们大家将更糟。

这个难题有两个解决方法：政治的解决方法和心理的解决方法。就那些心理的解决方法而言，最重要的是道德的解决方法。正如我所论证的，我们需要道德的解决方法的事例有许多。

我表述了四类这种解决方法。这些方式是由四种动机驱动的：值得信赖、不愿做搭便车者、想要满足康德的检验标准，以及充分的利他主义。每一种道德的解决方法都有两个形式。当这些动机中的一种引导某人作出利他主义的选择的时候，这个人所做的对他而言可能要么更糟、要么不糟。这个区分产生一些深刻的问题。我将简单陈述我的论证所假定的是什么。按照有关自我利益的所有令人信服的理论，符合我们利益的东西部分取决于我们的动机或者愿望是什么。如果我们具有道德动机，该利他主义的选择对我们将更糟这一点因而并不一定为真。但是这可能为真。即便如此，我们可能仍然作出这一选择。

我在此是在否定四个主张。有些人说，没有人做他所相信的对他更糟的事情。这点已时常遭到驳斥。另一些人说，每一个人所做的就定义而言是对他来说最好的。用经济学家的术语来说，它将"效用最大化"。既然这只是一个定义，它不可能是假的。但是它在此是不相关的。它与符合一个人自己的长期自我利益的东西根本无关。又有一些人说，美德总会得到报偿。除非有来生，这点也已经被驳倒。还有另外一些人说，美德是其自身的报偿。按照目标列表理论，有道德和道德地行动可能是那些使我们的生活变得更好的东西之一。但是按照该

理论的那些令人信服的版本，可能会有这样一些情形，在这些情形中道德地行动总体上会对某个人更糟。合乎道德地行动可能使这个人过多地丧失掉其他一些使我们的生活变得更好的东西。

回到我自己的主张。许多囚犯的两难困境需要道德的解决方法。要取得这些解决方法，我们必须直接倾向于作出利他主义的选择。每一种道德的解决方法都有两种形式。一种形式是废止该两难困境。在这些事例中，因为我们具有某种道德的动机，所以，如果他作出利他主义的选择则对每一人将更糟这一点并不为真。但是在其他事例中这一点仍然为真。即便在这些事例中，我们可能作出这个选择。每一个人可能出于道德的理由做他知道将对他更糟的事情。

对这第二种形式我们常常需要道德的解决方法。称它们为克己的。它们解决了那个实际上的难题。结果会对每一个人更好。但是它们并未废止该两难困境。一个理论难题仍然存在。

该难题是这样的。我们可能具有作出利他主义选择的道德理由。但是如果他作出自利的选择对每一个人将会更好。道德与自利相冲突。当这些冲突发生的时候，合理的行动是什么？

按照自利论，利己的选择是合乎理性的。如果我们相信自利论，我们对克己的道德解决方法将感到矛盾。我们将认为，要达到这样的解决方法，我们必须全都非理性地行动。

许多学者反对这个结论。其中一些人主张说，那些道德理由并不比那些自利理由更弱一些。另一些则更大胆地主张说，它们更强一些。按照他们的观点，自利的选择是非理性的。

这场争论可能看起来无法解决。如何能够对这两种理由彼此加以权衡？道德理由当然在道德上是至高无上的。但是自利的理由以自利的标准衡量则是至高无上的。我们从何处可以寻获一个中立的尺度？

32.以其自身的标准衡量在囚犯的两难困境中自利论失败了吗？

一直有人主张说，我们无需一个中立的尺度。有一种见识是认为自利论在囚犯的两难困境中是自败的。已有人主张说，既然这是真的，那么即便以自利的标准衡量，道德理由也优于自利理由。

正如我们已经知道的，S可能是个体性地间接自败的。如果某人从不克己，对他而言或许更糟。但是在囚犯的两难困境中这一点并不为真。此处的糟糕结果是由行动而不是由意向所产生的。而且哪一个选择对每个人更好一些也是清楚的。就每一个人而言这一点是真的，那就是，如果他作出利他主义的选择，这肯定将对他更糟。S告诫每一个人要作出利己的选择。而且，无论他人做什么，如果他自己作出这一选择，那么对每一个人而言将会更好。S在此并非是个体性地自败的。但是在第22节所界定的那个意义上，S是直接地集体性地自败的。如果大家全都成功地遵循S，对每一个人而言这将比没有人遵循S更糟。

这表明，如果大家全都遵循S，我们就是非理性的吗？我们可以从更小的一个问题开始。如果我们相信S，甚至以其自身的标准衡量我们的理论，也会是失败的吗？

我们会回答说："不。每一个人对自利的追求对他自己会更好。它是成功的。S为什么是集体性地自败的？只是因为对自利的追求对他人会更糟。这并不影响它的成功。它并不是'行善'。"

如果我们是利己的，我们当然要探讨那些囚犯的两难困境。这些并非是古典经济学家所喜爱的事例，其中如果每一个人追求自利，每一个人获益。我们或许说："在这些事例中，S既发挥作用，又认可该境况。在囚犯的两难困境中，S仍然发挥作用。每一个人仍然将从其对自利的追求中获得利益。但是由于每一个人从他人的利己行动中所损失掉的要更多，S在这里谴责该境况。"

这看起来可能是一种遁词。倘若当我们全都追求自利的话，则对每一个人会更糟，那么看起来自利论可能应当谴责它自己。假设，在某个别的群体中，在面对同样的一些两难问题时，全都作出利他主义

的选择。他们或许对我们说："你们认为我们是非理性的。但是我们比你们处境更好一些。甚至以利己的标准衡量我们做的也更好一些。"

我们会回答说："那只是文字游戏。只是在你们处境更好一些的意义上说你们'做得更好'。以利己的标准衡量你们每一个人是在做更糟的事情。每一个人是在做他所知道的将对他更糟的事情。"我们可能补充说："对我们每一个人而言更糟的是，在我们的群体中没有什么傻瓜。你们每一个人有更好一些的运气。尽管你的非理性对你是糟糕的，但是你从他人的非理性中所获得的甚至更多。"

他们可能回答说："你部分是对的。以利己的标准衡量，我们每一个人是在做更糟的事情。但是，尽管每人是在做更糟的事情，但是我们是在做更好的事情。这不是一种文字游戏。我们每一个人是因为我们所做的而处境更好一些。"

这个提法令人更加有所期待。回到更简单一些的两人事例。每一个人要么会给自己带来利益（E），要么给他人某种更大的利益（A）。结果会如同下述所示。

		你	
		做 E	做 A
我	做 E	对每一个人来说，第三好	对我最好，对你最糟
	做 A	对我最糟，对你最好	对两者来说，第二好

为确保没有哪一种选择能够影响到另一种选择——可能会产生互惠——假设我们不能够交流。如果我做A而不是E，那么对我将更糟。无论你做什么都是如此。对你而言情况同样成立。如果我们两者都做A而不是E，每一个人从而是在做以自利的标准衡量更糟的事情。该提法是说我们是在做较好一些的事情。

使人更加有所期待的是它把每个人与我们加以对照。正如前文所见，对于每个人为假的事情，对我们则可能为真。例如，这样一点能

够为真，那就是，尽管就每个人而言，没伤害到他人，但是我们合起来伤害到他人。如果我们两者都做A而不是E，尽管以自利的标准衡量，每个人是在做更糟的事情，但是我们合起来是在做更好的事情，这一点为真吗?

我们可以运用这个测试。自利论给每个人设定某个目标。如果就对他来说可能的那些行动中他所做的是使他的S设目标更好地达到，那么每个人以S的标准衡量则做得更好。如果就对我们而言可能的那些行动中我们所做的是使每个人的S设的目标更好地达到，那么我们以S的标准衡量会做得更好。该检验标准看起来是公平的。它或许表明，如果每个人做他自己所能够做的、以S的标准衡量最好的事，我们合起来不会做得更好一些。

当我们在度量成功的时候，只有终极目标重要。假设我们正在努力抓挠我们的后背。每个人的终极目标或许是他不再感到痒痒。那么如果我们抓挠彼此的背部，我们会做得更好。但我们或许是些柔体杂技演员：每人的终极目标或许是他的背部被他自己抓挠到。如果我们抓挠彼此的背部，那么我们会做得更糟。

自利论为每个人设定的终极目标是什么？是他的利益得到推进，还是他的利益被他自己推进？按照自利论，如果某人的利益被他自己所推进，这个人是在合理地行动。从而我能够重新陈述我的问题。S为每人设定的终极目标是什么？是他的利益得到推进，还是他合理地行动?

在第3节中，我为下述答案作了辩护。像所有的有关合理性的理论一样，S为每人设定他合理地行动这一形式目标。但是按照S，这个形式目标本身不是一个实质性目标。S为每个人设定一个终极的实质性目标：他的生活对他而言过得尽可能好。按照关于自利的快乐论，合乎理性与合理地行动不是这个目标的组成部分。它们两者都是纯粹的手段。按照其他一些有关自我利益的理论，合乎理性与合理地行动并不是纯粹的手段。无论它们的后果可能是什么，它们两者都是S为

每人设定的终极的实质性目标的组成部分。但是当它们总体说来对某人会更糟的时候，这一点就不会是真的。

我们可以设想一个为每个人设定这样一个实质性目标的理论：他的利益被他自己所推进。某个相信这个理论的人或许粗糙地错解尼采，看重"最强烈地依靠自己"。[1]如果我们两者都做A而不是E，以这些亚尼采术语看来我们会是在做更糟的事情。每人的利益会得到更好的推进。但是没有哪一个人的利益会被他自己所推进，所以亚尼采目标的达到程度会更糟。

如果我们两者都做A而不是E，我们以S的标准衡量是在做得更好吗？我们使每人的利益得到更好的推进。在这个方面，我们以S的标准衡量是在做得更好一些，使每人的S设目标得到更好的实现。按照有关自利的快乐论，这圆满地回答了我的问题。按照这个理论，S宣称任何行动都是纯粹的手段。目标总是关于人的意识生活的后果。（尼采的"金发野兽"据说是狮子。但是对它们而言行动也纯粹是一种手段。它们更愿意吃其他狮子的猎获物。）

按照其他一些关于自利的理论，必须说的还有很多。根据S，如果我们两者都做A而不是E，那么我们两者都是在非理性地行动。每个人是在做他所知道的对他将更糟的事情。按照某些有关自利的理论，合乎理性与合理地行动是S为每个人所设定的目标的组成部分。按照这些理论，一些貌似的囚犯的两难不是真的两难。我在注解[2]中讨论这些情形。

33.对道德的另一个弱辩护

在那些真正的两难问题中，如果我们两者都做A而不是E，那么以S的标准衡量，我们是在做更好的事情。我们是在造成每个人的S设目标更好地达到。按照所有关于自我利益的理论这都是如此。如果我们做S告诫我们不要做的事情，那么我们以S的标准衡量会做得更好。

这表明S以其自身的标准衡量是失败的吗？看起来可能如此。而且把S与道德加以比较具有相当的诱惑力。我们或许说，“自利论滋生冲突，告诉每个人针对他人而行动。这就是为何如果每一个人都追求自利，这对每一个人而言可能是糟糕的。自利论分化之处，由道德加以统一。道德告诉我们要共同努力——尽最大努力去做我们所能够做的最好的事。甚至按照自利所提供的衡量尺度，道德也因此是胜者。此乃我们从囚犯的两难困境中所得悉的。如果我们不再是利己的，而是变成道德的，那么甚至以利己的标准衡量我们也会做得更好。”[3]

这论证无法成立。我们做得更好，但是每个人却做得更糟。如果我们两者都做A而不是E，我们使结果对每个人而言更好，但是每个人使结果对他自己而言则更糟。无论他人做什么，对每个人而言如果他做E会更好一些。在那些囚犯的两难困境中，问题就在于此。每个人应当尽其所能做对他自己而言最好的事情呢？还是我们应当尽我们之所能做对每个人最好的事情？如果每个人做对他自己最好的事情，我们做的就会比我们本会为每个人所做的更糟。但是只有当每个人做对自己而言比本会做的更糟的事情时，我们做的对每个人而言则更好。

这只是一个更为广泛的问题的一个特殊事例。考虑一下任何一个有关我们有理由做的事情的理论。或许有这样一些事例，如果每个人以该理论的标准衡量做得更好一些，那么我们则会变得更糟，反之亦然。称这样的一些事例为每人-我们的两难困境（Each-We Dilemmas）。一个理论即便并不关乎符合我们利益的事情，它也能够产生这样的一些两难。

后果论的诸理论不会产生这样的两难困境。正如我们在第21节中已经看到的，这是因为这些理论是行动者中性的，为所有的行动者设定共同的目标。

如果一个理论的确产生每人-我们的两难困境，那么，这说明什么这一点可能是并不明显的。再考虑一下自利论。它告诉每个人要做他所能够做的对他自己而言最好的事情。我们是在讨论这样一些事例，

就是，如果我们全都追求自利，我们是在做对每个人而言更糟的事情。自利论在此是直接地集体性地自败的。但是我们不能认为这是一个缺陷。S为什么应当是集体性成功的？S在个体层次上为什么还不够成功？

我们或许说："一个理论不能只适用于一个单一的个人。如果对我来说做对我最好的任何事情是合理的，那么对任何人来说做对他自己最好的任何事情就是合理的。任何一个具有可接受性的理论因而必定是在集体性的层次上是成功的。"

这包含着一种混乱。如果一个理论适用于每一个人则称之为普遍性的，如果它宣称在集体性的层次上是成功的则称之为集体性的。有些理论具有这两个特征。一个例子就是康德的道德论。它告诉每个人只做他会合理地意欲每一个人去做的事情。每个人的计划或者策略必须在集体层次上得到检验。对一个康德主义者而言，道德的本质是从每个人推及我们。

在集体层次上——就像对"我们应当如何行动"这个问题的回答——自利论会宣告自己无用。假设我们是在选择什么样的道德规范将得到公开的鼓励，受到公开鼓励并在学校中得到灌输。S在此处会告诉我们投票反对它自己。如果我们是在选择一个集体性的规范，那么利己选择会是道德的某种版本。

S是普遍性的，适用于每一个人。但是S并非一个集体性的规范。它是一个关于个体合理性的理论。它所回答的是我上面所问的那个较小的问题。在那些囚犯的两难困境中，S是个体性地成功的。既然它只是集体性地自败的，S以它自己的标准衡量并不失败。S并不宣告自己无用。

34.时际两难

许多糟糕的理论并不宣告自己无用。所以那个更大的问题仍然悬而未决。在这样的一些事例中，去做什么才是合理的？

引入另一个常见的理论可能有所帮助。该理论告诉每个人要做将最好地达到他当前目标的事情。称这个理论为当前目标论（缩写为P）。假设在某个囚犯的两难困境中，我的目标是结果对我而言最好。根据P，那么，它是合理的益己的选择。如果我的目标是给他人带来利益，或者通过康德的检验标准，那么合理的是利他主义的选择。如果我的目标是去做他人所做的——大概因为我并不愿意做一个搭便车者——那么，哪一个选择才是合理的并不明确。这有赖于我关于他人做什么的信念。

正如这些评述所表明的，P可能与S相冲突。那种会最好地达到我目前目标的东西可能有悖于我的长期利益。由于这两个理论可能冲突，S的捍卫者必定拒斥P。

他们可能指出，即使在个体的水平上，P也会是直接自败的。它会产生时际两难。对于那些较少在乎它们更遥远的未来的人们来说，这些将是最常见的两难。假设我就是这样的一个人，而且在不同的时间我具有不同的目标。在每一时刻，我要么会（1）做将最好地达到我当前目标的事情，要么会（2）做将最好地达到或者随着时间的推移使我能够达到我的所有目标的事情。根据P，我应当总是做（1）而不是（2）。只有如此，我在每一时刻做我所能够做的、以P的标准衡量最好的事情。但是随着时间的推移P以同样的标准衡量我则可能做得更糟。随着时间的推移，我可能较少成功地达到我在每一时刻的目标。

这里举一个微不足道但就我的情形而言为真的一个例子。在每一时刻，如果我不把精力浪费在保持整洁上，我将最好地达到我当前目标。但是如果我永不整洁，这可能造成我在每一随后的时刻较少地达到。而且我的不整洁可能使我最初所试图达到的东西落空。那么，尽管令人可惜，但这样一点将是真的，就是从不保持整洁使我在每一时刻较少达到目标。

35.对自利论的一个弱辩护

那些相信自利论的人们可能诉诸这样一些情形。他们或许说："当前目标论在此是直接自败的。甚至以P的标准衡量，S也更优越一些。利己的选择是（2）。如果你总是做（2）而不是（1），在每一时刻你将更加有效地达到你的那些目标。如果你遵循S，那么甚至以P的标准衡量你做得也更好一些。"

就像对道德的那个类似的辩护一样，这个论证是失败的。如果我遵循S，随着时间的推移，我做得更好。但是在每一时刻我则做得更糟。如果我总是在做（2），那么在每一时刻我是在做将较少有效地达到我当时所具有的那些目标的事。（1）是将最好地达到这些目标的。

这一区分可能难以把握，假设我总是做（1）而不是（2）。那么这一点将是真的，那就是，随着时间的推移，我将较少有效地达到我在每一时刻所具有的目标。如果此点为真，在每一时刻我将更为有效地达到我在那个时刻的目标这一点如何可能也是真的？要看清这一点何以可能，我们可以回想一下人际两难。"我们"一词由"随着时间的推移"替代，而"每个人"一词由"每一时刻的我"替代。在人际两难中，仅当各自所做的对他自己而言比他本来所能做的要更糟的时候，我们所做的对各自而言才更好。在时际两难中，仅当我各次所做的比随之本来所能做的要更糟的时候，我才随着时间的推移每次做得更好。

正如这些主张所提示的，每人-我们两难是一个甚至更为广泛的难题的一个特殊事例。称这些为理由-相对性两难。S产生每人-我们两难，因为其理由是相对于行动者的。根据S，我能够有理由去做而你有理由不去做的事情。P产生时际两难，因为其理由是时间上中立的。根据P，我现在能够有理由去做而我后来将有理由不去做的事情。

P能够是时际性地自败的。但是P并不主张自己在时际性的层次上是成功的。它是一个有关我们在每一时刻有理由去做什么的理论。甚

至在时际两难中，P也是在每一时刻成功的。如果我总是遵循P，做（1）而不是（2），在每一时刻我就是在做将最好地达到我在那个时刻的目标的事。既然P是有关我们在每一时刻有理由去做什么的理论，那么在此处它以它自己的标准衡量并不是失败的。P并不宣告自己无用。

自利论者必定主张，尽管如此，P也应当被拒斥。他可能说：“任何一个可接受的理论必定是时际性地成功的。P并不主张这样的成功这一点并不是一种辩护。这只是表明P带有结构上的缺陷。如果一个理论是时际性地自败的，这足以表明它应当遭到拒斥。”

这些主张可能无法支持S。如果P被它是时际性地自败的这个事实所驳倒，那么何以S不被它是人际性地或者集体性地自败的这个事实所驳倒？而且，如果S并不主张自己是集体性地成功的这一点是一个很好的回应的话，当前目标论者何以能不作类似的回应？

正如这些评述所表明的，自利论会从两个方向遭到挑战。这使它更加难以得到辩护。对两个挑战中任何一个挑战的回答都可能从根本上削弱对另一个的回答。

一个挑战来自道德理论。另一个挑战来自当前目标论。这个理论有数个版本。最简单的版本是工具论。根据这一理论，每一个人最有理由做的事情就是将会最好地达到他当前目标的任何事情。这个理论把行动者的目标当作设定的，只对手段进行讨论。没有什么目标被说成是非理性的。任何一个目标都能够为行动提供良好的理由。

P的另一个版本是审慎论。这一理论所诉诸的不是行动者当前的实际目标，而是他如果知道相关的事实而且思虑清楚的话，现在本会有的目标。根据这个理论，如果一个目标经受不住这样一个审慎的过程的话，那么它就并未向行动提供一个良好理由。

P的第三个版本又以另一种方式对目标进行批判。按照这个理论，即便它们能够经受住这样一个审慎的过程，某些种类的目标也是内在地非理性的，并不能够为行动提供良好的理由。每一个人最有理

由做的事就是任何将最好地达到他的那些并非非理性的当前目标的事。此乃批判性的当前目标论。

在其所有的版本中，P时常与自利论相冲突。某人可能知道那些事实而且思虑清楚，但却具有他明知有悖于其长期的自我利益的目标。而且我们会相信，这些目标中的一些并非是非理性的。可能可以举出一些例子：给他人带来利益、发现真理和创造美等。我们会得出结论认为，对这些目标的追求并不比对自我利益的追求更缺乏合理性。按照这一观点，即便当行动者明知他所做出的行动有悖于其长期的自我利益的时候，追求这样的目标也并非是非理性的。

自利论者必定拒斥这些主张。他必定坚持认为，行动的理由不可能是相对于时间的。他可能说："理由的力量随着时间的推移延伸开来。既然我将有理由促进我未来的目标，那么我就有理由现在这样做。"这个主张是自利论的核心所在。

许多道德论者还提出另一个主张。他们认为，某些理由并非是相对于行动者的。他们可能说："理由的力量可能不仅随着时间的推移延伸开来，而且在不同的生命中延伸开来。于是，如果我有理由解除我的疼痛，对你而言这也是个理由。你有理由解除我的疼痛。"

自利论者提出的是第一个主张，但是拒斥第二个主张。他可能发现，对这个立场的两边都加以辩护是困难的。在对道德论者的回应中，他可能问，"我为什么应当赋予那些并非是我的目标的目标以权重？"但是当前目标论者会问，"我现在为什么应当赋予现在并非是我的目标的目标以权重？"自利论者可能以诉诸时际两难予以回应，其中当前目标论是时际性地自败的。但是他可能遭到人际两难的挑战，其中他自己的理论是集体性地自败的。道德论者或许说："对自利论的论证把我们带到这个理论之外。恰当地理解，它是一个对道德的论证。"

在第二编中，我将紧跟这个思路。但是现在首先应当讨论别的。在人际层次上，对比并非是在自利论和道德之间。

36.常识道德如何是直接自败的

正如我在第22节中所蕴含的，自利论并非是能够产生每人-我们两难的惟一理论。这样的一些事例可能发生在（a）某个理论T是相对于行动者的——给予不同的行动者以不同的目标——的时候，（b）每一个人的T设目标的达到部分取决于他人所做的是什么的时候，以及（c）每一个人的所作所为将不影响到这些他人的所作所为的时候。这些条件对常识道德而言是经常成立的。

我们绝大多数人相信，存在着特定的一些人，我们对这些人具有特别的义务。我们与这些人处在特定的关系之中——诸如我们的子女、父母、朋友、恩人、学生、病人、客户、同事、工会会员、那些我们所代表的人们或者我们的同胞。我们相信，我们应该力求拯救这些人摆脱某种伤害，而且应该力求给予他们某种利益。常识道德主要寓于这样的义务之中。

履行这些义务优先于帮助陌生人。这一优先权不是绝对的。我不应为了避免我自己的孩子受伤，而不救一个陌生人的生命。但是我应该拯救我的孩子脱离某种伤害，而不是使一个陌生人避免稍大一点的伤害。由于我对我的孩子的义务，任何时候我都可以不理会我本可以在别处做稍大一些的善行这一点。

当我努力保护我的孩子的时候，我的目标应当是什么？只是孩子应当不受到伤害？还是由我救他脱离伤害？如果你本有更好的机会救他脱离伤害，那么我坚持由我救他就是错误的。这表明我的目标应当呈现为更简洁的形式。

我们可能面对父母两难。考虑一下

事例一。我们不能够交流。但是每个人要么会（1）救他自己的孩子脱离某种伤害，要么（2）救别人的孩子脱离另一个稍大一点的伤害。结果如下所示。

		你	
		做（1）	做（2）
我	做（1）	我们的两个孩子都受到那个较大的伤害	我的孩子没有受到任何伤害，你的孩子遭受到所有两个伤害
	做（2）	我的孩子遭受到所有两个伤害，你的孩子没有遭受到任何一个伤害	两个孩子遭受到那个较轻的伤害

既然我们不能够交流，两个人的选择中没有哪一个将影响到另一个的选择。如果每个人的目标应当是他的孩子不遭受伤害，那么每个人在此应当做（1）而不是（2）。于是每个人会确保他的孩子受到的伤害较轻一些。无论另一个人做什么这一点都是如此。但是如果两者都做（1）而不是（2），我们的两个孩子就都会受到较重的伤害。

接着考虑一下我应该力求给予我的孩子的那些利益。我的目标在此应当是什么？即便对我的孩子而言会更糟，我应当坚持由我给我的孩子带来利益吗？一些人总会给出否定的回答。但是这一回答可能过于笼统了。它把父母的关爱当作纯粹手段来看待。我们可能认为父母的关爱远不止于此。我们可能同意，就某些种类的利益而言，我的目标应当呈现简单一些的形式。它应当只是在结果上对我的孩子来说更好一些。但是可能还有其他一些种类的利益，我的孩子应当从我这里得到。

就这两种利益来说，我们可能面对父母两难。考虑一下

事例二。我们不能够交流。但是每个人要么会（1）给他自己的孩子带来利益，要么会（2）给他人的孩子带来略多一些的利益。结果如下所示。

		你 做（1）	你 做（2）
我	做（1）	对我们的两个孩子来说第三好的结果	对我的孩子来说最好，对你的孩子来说最糟
我	做（2）	对我的孩子来说最糟，对你的孩子来说最好	对两个孩子来说第二好

如果我的目标在此应当是对我的孩子来说有更好的结果，那么我应当再次做（1）而不是（2）。而且对你来说情况同样如此。但是如果我们两者都做（1）而不是（2），这对我们的两个孩子来说都将更糟。比较一下

事例三。我们不能够交流。但是我要么（1）能够使自己给我的孩子某种利益，要么（2）能够使你给你的孩子略多一些的利益。至于我，你可以作出同样的选择。结果如下所示。

		你 做（1）	你 做（2）
我	做（1）	每个人能够给他的孩子某种利益	我能够给我的孩子带来的利益最大，你能够给你的孩子带来的利益最小
我	做（2）	我能够给我的孩子带来的利益最小，你能够给你的孩子带来的利益最大	每个人都能够给他的孩子带来更大的利益

如果我的目标在此应当是我给我的孩子带来利益，那么我再次应当做（1）而不是（2）。而且对你来说情况同样如此。但是如果我们两者都做（1）而不是（2），那么每个人能够给他的孩子带来的利益较少。请注意这两个事例之间的差异。在事例二中，我们关心的是发生什么。每个人的目标是结果对他的孩子来说更好一些。这是另一个人能够直接导致它实现的一种目标。在事例三中，我们所关心的是我

们做什么。由于我的目标是我给我的孩子带来利益，所以你不能够代表我做到这一点。但是你能够帮助我做到这一点。那么你能够间接地帮助我的目标得到实现。

两人的父母两难不太可能发生。但是我们时常面对多人的两难困境。这一点通常是真的，就是，如果全都而不是没人给予我们自己的子女优先权，这要么将对我们的所有子女而言更糟，要么将可能使每人给他自己的孩子带来的利益较少。于是有许多公共的善：即无论我们是否有助于它们的产生，都会有益于我们子女的那些结果。就每一个为人父母的人而言，这样一种情况能够为真，那就是，如果他不提供帮助，这对他自己的那些子女而言将会更好。他所节省下来的东西——无论是金钱、时间还是能源——他都能够用来只给他的子女们谋利益。但是如果没有为人父母者帮助生产这样的公共的善，这对我们的所有子女而言将比所有的人全都这样做要更糟。在另一个常见的事例中，就像渔民两难中那样，每个人要么（1）增加自己的所得，要么（2）（凭借自制）增加别人的所得。在此就每个人而言这一点将是真的，那就是，如果他做（1）而不是（2），他能够给他的子女们带来更多的利益。无论其他人做什么这一点都是如此。但是如果所有渔民全都做（1）而不是（2），那么每人能够给子女带来的利益则要少一些。这些只是这样的一些事例可能出现的那些方式中的两个而已。别的方式还有许多。

类似的评述适用于所有类似的义务——诸如那些对学生、对病人、对客户或者对百姓的义务等。就所有这样的义务来说，存在着无数就像我所举出的三个父母两难那样的多人版本。它们就像多个囚犯的两难困境一样常见、一样变化多端。如同我们刚刚所见的，它们时常具有同一原因。

下面是又一个以此为真的方式。假设在原先的囚犯的两难困境中，必须作出抉择的是我们的律师。这产生出囚犯律师的两难。如果两个律师都给予他们自己的委托人以优先权，这对两个委托人而言将

比没有哪一个律师这样做的话更糟。任何自利的两难于是都可能产生一种道德上的两难。如果一个群体面对前者，另一群体可能从而面对后者。如果第二个群体中的每一个成员应该给予第一个群体中的一些成员以优先权的时候，情况就可能如此。当不同群体，诸如国家，面对一个自利两难的时候，类似的主张同样适用。绝大多数政府认为，他们应该给予他们的公民的利益以优先权。存在着数种方式，其中如果所有的政府全都给予而不是没有给予它们自己的公民以优先权，那么对它们的所有公民而言将会更糟。该难题出在给予优先权。无论是给予一个人自己还是给予他人，这都没有什么差别。

我所举出的例子全都涉及到伤害或利益。对常识道德的其他方面而言，该难题也可能出现。每当这样的道德给予不同的人们不同的义务的时候，它就可能呈现。假设每人要么会（1）履行他自己的一些义务，要么（2）能够使他人履行他们的更多义务。如果全都而不是没人给予他们自己的义务以优先权，那么，每人所能够履行的义务可能会更少一些。义务论者也可能面对每人-我们两难。

37.道德理论的五个部分

这些事例表明什么？我们绝大多数人接受我所称的常识道德的那种理论的某个版本。根据这一理论，存在着我们每人都应该力求达到的一些特定东西。这些就是我所称的我们的道德目标。当每人所做的是他可能做的那些行动中最好地达到其道德目标的行动，那么我们就成功地遵循了该道德理论。在我所举出的那些事例中，如果所有的人而不是没人成功地遵循该理论，我们肯定会使每个人的道德目标较差地达到。我们的道德理论在此是直接地集体性自败的。这是一个诘难吗？

让我们从一个更小一些的问题开始。我们能够修正我们的理论，以便它不会是自败的吗？如果没有这样的修正，我们的理论可能是最

好的可能的理论。我们首先应当认准我们理论中自败的是哪一个部分。

以完全依从为假定的道德理论的一个部分可能涵盖成功的行动。称道德理论的这一部分为理想行动理论。这一部分所说的是，假定我们全都努力而且全都成功，我们全都应当努力做什么。称之为我们全都应当理想地去做的。“全都”在此并不是指“每一个活着的人”。它意指“某个群体的成员”。

正如我在第一章中所论证的，决定什么是我们全都应当理想地去做的还并不够。我们必须把下面的四个事实考虑进去：

(a) 我们对我们行动的后果是什么通常并不确定；
(b) 我们中的一些人可能错误地行动；
(c) 我们的行动并非是我们的动机的惟一后果；
(d) 当我们感到懊悔或者责怪别人的时候，这可能影响到我们以后的行动，而且具有其他后果。

我们的道德理论因而可能具有下述五个部分。

	全都	每人
成功的行动	我们的理想行动理论，说的是当我们知道将成功的时候我们全都应当理想地去做什么	我们的实践行动理论，说的是鉴于(a)和(b)我们中间的每人应该做什么
动机	我们的理想动机理论，说的是鉴于(a)和(c)我们全都应当理想地具有的是什么动机	我们的实践动机理论，说的是鉴于(a)、(b)和(c)我们每个人应当有什么动机
责难和懊悔		我们的反应理论，说的是鉴于(a)、(b)、(c)和(d)，每个人应当为之受到责难和应当感到懊悔的行动是哪些

当我们在决定我们相信什么的时候，我们首先应当考虑我们的理想行动理论。问我们全都应当理想地去做的是什么，我们是在问我们的终极道德目标应当是什么。这些是我们的道德理论的基础。我们的理论的其他部分是：鉴于我们的终极道德目标，当我们考虑上述的四个事实的时候，需要主张的是什么。

38.我们如何能够修正常识道德以使其不会是自败的

假设我们接受某个版本的常识道德。在我所举的例子中，这一点为真：如果我们全都成功地遵循我们的道德理论，它将是直接地自败的。自败的是我们的理想行动理论。如果我们应当修正我们的理论的话，这是我们首先应当予以修正的部分。

称我们的理论为M，其修正版本为R。R的主张之一是

（R1）当M是自败的时候，我们全都应当理想地去做将造成每个人的M设目标更好地实现的事情。

于是在我举的所有的父母两难事例中，我们全都应当理想地去做（2）而不是（1）。这会使结果对我们所有的子女来说更好，而且能够使每个人给他自己的子女带来更多的利益。

（R1）修正的是我们的理想行动理论。如果我们修正我们的这一部分理论，我们将被自然地引导到修正其他部分。

首先考虑我们的实践行动理论。这一理论表述的是，鉴于（a）我们通常并不知道我们行动的后果是什么这一事实，以及（b）我们中一些人将会错误地行动这一事实，我们每个人应该做什么。

回到那个会给我们的子女带来利益的一个公共善的事例。保护稀缺资源就是这样的一个善。假设我们是渔民两难中贫穷的渔民，努力捕获足够的鱼以养育子女。因为过度捕捞，出现了渔业资源的下降。就每个人而言这一点是真的，那就是，如果他不节制自己的捕捞的

话，这对他自己的子女而言将稍好一些。他们将得到稍好一点的生活。无论他人做什么这一点都是如此。但是如果我们中没有人节制他的捕捞，这对我们的所有子女而言将比我们全都节制我们的捕捞的话要糟糕得多。我们所有的子女将得到差得多的生活。根据（R1），我们全都应当理想地节制我们的捕捞。如果一些人未能做到这一点，（R1）就不再适用。但是提出这样一个进一步的主张会是顺理成章的：倘若足够的他人也这么做的话，那么每个人都应当节制他的捕捞。

怎样才算足够？对这个问题存在着一个自然答案。考虑一下任何一个会给我们的子女带来利益的公共的善，而且仅当存在自愿贡献的情况下，才将提供这样的善。为简化起见，假定不存在这样一个上限：超过这个上限贡献就会被浪费。如果我们全都做贡献的话，我们的子女将得到最大的利益。假设我们每个人都知道将有某些父母不做贡献。那么必定存在某个这样的最小的数目k，如果k或者更多的父母做出贡献的话，这对每一个贡献者的子女而言将比没有人做贡献要更好。如果只有一个人做贡献，这对他的子女而言要比他不做贡献更糟。如果全都做贡献，这对我们所有的子女而言将比没有人做贡献更好。在一个人和全都之间的某处必定存在数目k，从更糟到更好的变化就是由此而来。[4]

数目k具有两个特别特征：（1）如果k或者更多的人做贡献，每一个贡献者就参与到一个策划之中，其净后果是给他自己的子女带来利益。每个贡献者的子女们得到的利益将比倘若没有人做出贡献的情况下要更多。（2）如果少于k数目的人做贡献，任何一个贡献者的子女们得到的利益将比倘若没有人做出贡献的情况下更少。（1）和（2）使k成为一个令人信服的道德临界值，在其之上每一个为人父母的人都应该做出贡献。我们能够主张

（R2）在这样一些事例中，如果他相信至少将有k数目的贡献者，每个人就应当作出贡献。

如果我们的实践行动理论主张（R2），我们道德观点中的这一变化通常可能改变我们所做的。

可能有这样的说法："既然（1）为真，那么我们无需主张说，如果他相信至少将有k数目的贡献者，每个人应该作出贡献。既然每个人所要加入的策划的净后果是给他的子女带来利益，那么他对子女的爱将使他想要加入这样的一个策划之中。这样做对他的子女将更好一些。"

这些主张是假的。假设至少k数目的父母做出贡献。这些贡献者们的子女们得到的利益将比没有人作出贡献的情况下更多。但是每一个贡献者却是在做对他自己的子女更糟的事情。如果每一个贡献者并未作出贡献，而是把他所节省下的一切——时间、金钱或者能源——用来只给他的子女们带来利益的话，对每一个贡献者的子女们而言本会更好。无论作出贡献的人还有多少，这一点都为真。既然每一个贡献者是在做对他自己的子女更糟的事情，那么我们需要主张的是，如果他相信至少将有k数目的贡献者，则每个人应该作出贡献。每个人应该作出贡献，因为，尽管每个人是在做对他自己的子女而言更糟的事情，但是k数目的贡献者合起来却是在做对所有他们的子女更好的事情。

要支持（2），我们还可以指出的是，如果任何一个为人父母者当他人作出贡献的时候自己并不作出贡献，那么他的子女们就将是些"搭便车者"。他们将以牺牲贡献者们的子女的利益为代价从公共的善之中获益。他们之所以将以牺牲贡献者们的子女的利益为代价而获益，是因为（a）他们将比贡献者们的子女得到更多的利益，而且（b）这之所以为真，是因为每一个贡献者所做的是对他自己的子女们而言更糟的事情。

类似的主张适用于我们其他的特别义务。根据常识道德，我们应该给予那些我们与之具有特定关系的人们的利益以优先权。除了我们的子女之外，还包括诸如这样一些人：父母、学生、客户、我们所代

表的那些人们和我们的同胞。姑且说我们与这些人有M关系。还有数种别的M关系。这些关系的共同之处在于，根据常识道德或者M，我们对所有那些我们与之有M关系的人们具有特殊的义务。

有无数这样的事例，就是如果每人给予他的M关系中的人们以优先权，这对所有这些人而言将比没有人给予这些人以优先权的情况下更糟。根据（R1），在这些情形中，我们全都应当理想地去做的事情是不给我们的M关系中的人们以什么优先权。如果我们遵循（R1）的话，这对所有这些人而言本会更好一些。

假设我们知道我们中的一些人将给予与我们有M关系的人们以优先权。（R1）不再适用。但是将再次仍然有某个这样的最小数目k，如果k或者更多的人不给予他们的那些与之有M关系的人们以优先权，那么甚至对这些人们而言这也将比所有的人给予他们的那些与之有M关系的人以优先权的话要更好一些。因而我们能够令人信服地主张

（R3）当M是自败的时候，如果相信至少有k数目的他人将以同样的方式行动，我们每人就应该不给那些与之有M关系的人们以优先权。

如果我们接受（R3），这通常可能再次改变我们所做的。如前所述，如果k数目的人们中的每一个都不给予与他有M关系的人们以优先权，每个人就是在做对这些人们而言更糟的事情。这就是为什么，根据常识道德，每个人都是在错误地行动的原因。但是尽管每个人是在做对与他有M关系的人们而言更糟的事情，但是这些k数目的人们合起来却是在做对所有这些人更好的事情。尽管这些k数目的人们是以一种M宣称为错误的方式行动，但是他们合起来导致每个人的M设目标更好地得到实现。既然他们遵循（R3）而不是M，甚至以M的标准衡量，他们合起来也做得更好。

其次考虑一下我们的理论中主张我们的动机应该是什么的那些部分。假设每个人要么会（1）救他自己的孩子脱离某种较轻的伤害，要么会（2）救另一个人的孩子脱离某种更大的伤害。根据（R1），

我们全都应当理想地做（2）。但是我们应当倾向于做（2）吗？如果那些较轻的伤害本身会是巨大的，那么这样一个意向或许与我们对子女的爱是不相容的。这可能引导我们决定应当保持倾向于做（1）。如果我们保持这样的倾向，我们可能在这样的情形中因而做（1）而不做（2）。那么我们的子女们会遭受更大的伤害。但是，如果我们要爱他们，这是他们必须付出的代价。

值得对极端事例作一表述。假设你和我每人有4个子女，他们全都处于生命危险之中。我们彼此是陌生人，而且不能够交流。每人要么会（1）救出他自己的一个孩子脱离危险，要么会（2）救出另一人的3个子女。如果我爱我的子女，我可能发现，让我以自己的1个子女死去为代价救出你的3个子女是不可能的。而且对你而言同样的情况也为真。那么我们两人将都做（1）而不做（2）。因为我们爱自己的子女，当我们本来能救出6个孩子的时候，我们却只救出他们之中的2个。

如果我们继续爱我们的子女，那么总体上说来将会更好。然而这可能有时使我们做（R1）和（R2）主张我们所应该做的事情在因果上是不可能的。但是也有许多别的事例，其中这一点不会为真。于是既爱我们的子女又对绝大多数公共的善做出贡献是可能的。

如果我们回到我们其他的一些特殊义务，那么主张我们倾向于不做（R3）要求我们所做的事情则更不能令人信服。于是，当对所有的公民而言更好的时候，不同国家的政府应该能够不给予它们自己的公民以优先权。当我们考虑具有不同的意向所带来的后果的时候，在绝大多数情形中令人信服的观点是，我们应当倾向于按照（R3）去行动。

我业已提出主张说，如果我们应该修正常识道德，那么我们就应该接受（R1）至（R3）这些主张。既然我们应该爱我们的子女，那么在某些特定的极端情形中我们不应该倾向于按照这些主张来行动。而

且可能还有别的特定的类似例外。但是在绝大多数情形中，我们应该倾向于按照这些主张行动。因而我们应当时常改变我们所做的。

39.我们为什么应该修正常识道德

如果我们修正“常识道德”（或者说M），那么我们就应该接受R的三个主张，即（R1）至（R3）。我现在回到那个主要问题。如果我们接受M，我们应该修正我们的观点吗？我们应该从M转移到R吗？在许多情况下常识道德是自败的这一点对它是一个诘难吗？如果是的话，R显然是补救良方。R修正M的自败之处。而且惟一的差异在于R并非是自败的。

首先记住，在这些情形中M是直接地自败的。问题不在于我们遵循M的尝试不晓得什么缘故失败了。那只会使M间接地自败而已。正如我已经论证的，这或许不是对我们的理论的什么诘难。此处的问题要严重得多。在我所表述的那些事例中，我们全都成功地遵循M。每个人成功地做了对他而言可能的那些行动中最好地实现了他的M设目标的事情。但是，因为我们全都成功地遵循M，我们导致每个人的M设目标实现得更糟。这就是使M自败之所在。能够主张说这不是什么诘难吗？这一点看起来相当值得怀疑。如果有任何一种假设，照此假设道德理论不应当是自败的这一点最为清楚的话，那就是得到普遍地成功遵循的假设。

其次记住，我所说的“目标”指的是实质性的目标。我已经忽略了我们的形式上的目标：避免做错。这可能看起来消除了该诘难。以那样一些事例为例，其中如果我们遵循M，要么结果对所有我们的子女而言将更糟，要么每个人能够给他的子女所带来的利益较少。我们或许说：“这些结果当然是不幸的。但是我们如何能够避免它们？只有通过不给予我们自己的子女以优先权。这可能会是错的。所以这些事例并未给我们的道德理论投下一层怀疑的阴影。即使要达到我们的其他道德目标，我们也决不应当错误地行动。”

这些评述有些无的放矢。的确，在这些事例中M形式上不是自败的。如果我们遵循M，我们并不是在做我们认为是错的事情。相反，既然我们相信M，那么我们相信不遵循M是错误的。但M是实质性地自败的。除非我们全都做我们现在认为是错的事情，否则我们将导致每个人的M设目标实现得更糟。问题是：这表明我们可能弄错了吗？还是我们或许应该去做我们现在认为是错的事情？我们不能够回答说，“不——我们决不应当错误地行动。”如果我们错了，我们就不会是在错误地行动。我们也不能够只是说，“但是，甚至在这些情形中，我们也应该给予我们自己的子女以优先权。”这恰恰假定我们并不是错误的。要辩护我们的理论，我们必须主张比这更多的东西。我们必须主张，在这样的事例中它是直接地、实质性地自败的这一点并不是对我们的理论的什么诘难。

如果我们的M设目标是否达到是无关紧要的话，这本不会是什么诘难。但这一点却的确是重要的。它重要的意义可能需要加以解释。如果我们尚没有错误地行动，在道德上它可能并不重要。但是它在一个具有道德蕴含的意义上是重要的。我们为什么应当努力达到我们的M设目标？部分原因在于，在这种另外的意义上，它们的达到是重要的。如果达到我们的道德目标并不重要的话，它们就会像一些毫无意义的规则了，只是要检验我们的服从性而已。

现在可能有这样的说法：“你称M为自败的。所以你的诘难必定诉诸M。你不应当诉诸某个对立的理论。这却是你刚刚所做的。当你主张说，我们的M设目标是否达到是重要的时候，你只是在主张说，如果它们不是的话，结果将更糟。这假定了后果论。以待证假定为论据来进行辩论。”

情况并非如此。在解释为什么的过程中，我将再次把两个差异结合起来。当我们的目标共有的时候，它们是对行动者中立的。其他目标则是相对于行动者的。任何目标可能要么涉及发生什么，要么涉及我们做什么。这带给我们4类目标。4个例子如下所示。

	涉及	
	发生什么	我们做什么
对行动者中立的	子女们 不挨饿	子女们得到 他们自己父母的关照
相对于行动者的	我的子女们 不挨饿	我关照 我的子女们

当我主张说，我们的M设目标是否实现是重要的时候，我不是在假定只有结果重要。我们的一些M设目标涉及的是我们做什么。于是父母的关爱对我们来说可能不是一个纯粹的手段。我也不是在假定行动者中性论。既然常识道德绝大部分是相对于行动者的，将假定作为论据来进行辩论。但这并不是我在做的。

此处要点有二。首先，我不是在假定重要的是M设目标的实现。假设我要么会（1）促进我自己的M设目标，要么会（2）更有效地促进你的M设目标。根据M，我在此处应当做（1）而不是（2）。我从而会造成M设目标总体上实现得更糟。但是这并不使M成为自败的。如果我遵循M，我会造成我的M设目标实现得更好。在我所举出的事例中，问题的关键不在于，如果我们全都做（1）而不是（2），那么我们会造成M设目标实现得更糟。要点在于我会造成我自己的每一个M设目标实现得更糟。我们不仅以行动者中性的标准衡量做得更糟，而且以相对于行动者的标准衡量也是做得更糟。

第二个要点是，这在一种相对于行动者的意义上可能是重要的。追忆一下自利论将有所帮助。在囚犯的两难困境中，这个理论是直接自败的。如果所有人而不是没有人成功地遵循S，我们将从而使每个人的S设目标实现得更糟。我们将使结果对每一个人而言更糟。如果我们相信S，我们将会认为这重要吗？抑或只有是否每一个人实现他的形式上的目标——避免非理性——才重要？答案是明显的。S为每一个人设定了结果对他自己而言要尽可能地好这样一个实质性的目标。该目标的实现是重要的。而且是在一种相对于行动者的意义上重要

的。如果我们相信S，我们将相信，在那些囚犯的两难困境中，如果我们全都遵循S则对我们每一个人而言更糟这一点是重要的。尽管它们并未驳倒S，但是以自利的标准衡量，这些事例则是令人遗憾的。在主张这一点上，我们无需诉诸S的行动者中性的形式——功利主义。自利论是关于合理性的而不是关于道德性的。但是对比表明，在讨论常识道德时，我们无需将假定作为论据来进行辩论。如果我们的M设目标得到实现是重要的，那么这在一种相对于行动者的意义上可能是重要的。

这重要吗？请注意，我不是在问这是否是全都重要。我不是在提议说，我们的形式上的目标的实现——避免做错——是一种纯粹的手段。尽管为一些后果论者所假定，但这并不是我们绝大多数人所相信的。我们甚至可能认为，避免做错总是最重要的。但是在此这并不相关。我们在问实质性地自败是否给M投下一层怀疑的阴影。这能表明在这样的一些事例中M是不正确的吗？最重要的是避免做错这一点可能是真的。但是这一点的真实性并不能够表明M是正确的。它不能够帮助我们决定什么是错误的。

我们能够主张说，避免做错是全都重要吗？如果是这样的话，我的那些例子就会什么也没表明。我们会说："实质性地自败在常识道德这一事例中不是损害性地自败的。"我们能够以这一方式捍卫我们的道德理论吗？就一些M设目标的事例而言，我们也许能够。考虑一下一些微不足道的承诺。我们可能既认为我们应当努力遵守承诺，又认为，如果不是出于我们自己的过失而没有履行这些承诺并不会是重要的。但是我们并不是对我们所有的M设目标都抱有这样的信念。如果我们的子女遭受伤害而痛苦，或者我们能够给予他们的利益要更少，那么这是重要的。我们的道德并非一套只是意在检验我们的服从性的无意义的规则。

最后请记住，在我所举的例子中，M是集体性地但并非个体性地自败的。这会提供一种辩护吗？这是我所提出的中心问题。正是因为

M是个体性地成功的，在集体性的水平上它能够是直接地自败的。如果我们全都做（1）而不是（2），则我们成功地遵循M这一点为什么为真？因为每个人是在做对于他而言可能的那些行动中最好地达到他的M设目标的事情。我们从而使每个人的M设目标达到得更糟这一点大概不是什么诘难吧？

谨记自利论将再次有所帮助。在那些囚犯的两难困境中，S是集体性地自败的。如果我们是在选择一种集体性的规范、某种我们全都要遵循的东西的话，那么S就会告诉我们拒斥它自己。自利的选择会是道德论的某个版本。但是那些相信S的人们可能认为这是不相关的。他们会说："自利论并不宣称是一种集体性的规范。它是一种个体合理性理论。就S的情形而言，集体性地自败的并不是损害性地自败的。"

我们能够以这种方式捍卫常识道德吗？这取决于我们有关道德本性和道德推理的观点。按照绝大多数观点，答案会是否定的。按照这些观点，道德本质上是一种集体性的规范——对"我们全都应当如何行动？"这一问题的回答。对该问题的一个可以接受的回答必定是在集体性水平上是成功的。该问题的答案不可能是直接集体性地自败的。如果我们相信常识道德，那么我们从而应当修正这一理论，以便其不至于在这个方面是自败的。我们应当接受R。

首先来考虑康德有关道德推理本质的观点。假定我现在面对父母两难中的一个两难。当对每一个人的子女而言，包括我的子女在内，这会更糟的时候，我会合理地希望所有人全都给予他们自己的子女以优先权吗？答案是否定的。对康德主义者而言，道德的本质是从每个人转移到我们。每个人只应当做他能够合理地希望我们全都做的事情。康德式的道德不能够是直接集体性自败的。

有数位学者接受有关道德推理本质的一种康德式观点，但是也接受某个形式的常识道德。如果他们保持他们的康德式的观点，这些学者们应该转移到修正版本的R。

另一些学者持有关于道德本质的建构主义者观点。对他们而言，道德是社会创造出来的某种东西，或者是社会成员们会合理地认可为指导他们行动的东西。这是另一种有关具有可接受性的道德理论不能够是直接集体性自败的观点。那些持有这样一种观点的人不能够继续接受常识道德的某个版本。他们必须移动到R的相应版本。

那些持有一种建构主义者观点的人可能质疑我有关道德理论的划分。（R1）修正我所称的我们的理想行动理论。建构主义者们可能认为一个道德理论无需这个部分。但是他们不能够反对我的这样一个提法，那就是，假定我们全都将努力并且取得成功，我们应当问我们全都应当做什么。对这个问题的回答最糟的是它是多余的。如果一个建构主义者问我们全都理想地做的是什么，他的回答不可能是常识道德的某个版本。如果他接受这个道德的某个版本，他必须转移到（R1）的相应版本，即他的道德的那个修正版本，它是不会直接集体性自败的。而且，既然他应当接受（R1），那么他也应当接受（R2）和（R3）。他应当修正他的实践行动理论，即曾经是其整个理论的那一部分。

按照有关道德推理本质的其他观点中的绝大多数，道德本质上是一个集体性规范。[5]按照这些观点，如果我们接受常识道德，我们必须移动到R。但是一些相信常识道德的人们可能没有什么关于道德推理本质的观点。这些人们本会主张说，即便存在着他们的道德直接自败的许多事例，但这并非对他们的道德的什么诘难，而且也不是什么移动到R的根据吗?

这样的主张是不具有可信度的。而且值得提一下常识道德如何接近于告诉我们要转移到R。假设在我所面对的一个父母两难中，我们全都能够顺利地进行交流。那么我们将得到常识道德这样的告诫，就是要作出我们全都遵循的一个有条件的联合承诺，但不是该种道德，而是我的修正了的R。如果我与其他人加入到这个有条件的联合承诺之中，这对我的子女而言将更好。如果我与其他每一个人都有条件地

承诺没有人给予他们自己的子女以优先权，那么我对我的子女的特殊义务从而将得到最好的履行。作出这一有条件承诺将是我能为自己的子女所做的最好的事情。如果我承诺遵循R，条件是其他每一个人也都作出相同的承诺，那么其他人只是因为我作出了这个承诺才将遵循R。如果他们遵循R，对我的子女而言将会更好。所以我的承诺使结果对我的子女而言更好。

类似的评述适用于常识道德自败的所有其他事例。这是一些这样的事例，其中我们认为应该给予我们其他的与之有M关系的人们以优先权，诸如我们的父母、学生、病人、客户，或者我们所代表的那些人们等等。在所有这些事例中，如果我们能够顺利地沟通，我们将得到常识道德这样的告诫，就是，作出我们将全都遵循R这个有条件的联合承诺。那么，因为我们所承诺的，这会成为每人应该做的。我们在此不会摒弃常识道德。我们会运用该道德的部分去改换我们所应该做的东西的内容。

下面假设在父母两难中，我们不能够沟通。那么我们将不能达到这个"道德解决方法"。常识道德现在告诉每人给予他自己的子女以优先权。对我们的所有子女而言这将会更糟。类似的评述适用于其他情形。因为我们不能够交流，而且因而不能够作出那一联合承诺，所以我们的道德不可能告诉我们遵循R。如果我们本可交流，并且评论一下"我们全都应当做什么"这一问题的话，那么我们的道德本会告诉我们要承诺不给予我们自己的子女、父母、学生、病人等等以优先权。如果我们不能够交流的话，我们的道德本会告诉我们要承诺不做它所告诉我们去做的。显然，按照我们的道德，如果我们本可交流并且承诺遵循R的话会更好。这点提供了一个我们的道德本身告诉我们去接受它本身的这个修正版本的道理。[6]

还有一个进一步的根据支持我们应该修正常识道德的想法。这个道德理论犯了我所称的道德数学中的第二个错误。它忽略了行动集合的后果——我们合起来所做的事情的后果。第三章表明了这是一个错

误。而且，在表明这一点的时候，我不是在假定后果论。那些拒斥C的人们会同意，在我所举的一些例子中，我们不应当忽略我们合起来所做的事情的后果。

每当常识道德直接集体性自败的时候，它都忽略这些后果。它告诉每个人去做将最好地实现其M设目标的事情。这个主张假定，考虑每一个人所做的事情的后果就足够了。在那些事例中，如果每个人做最好地实现其M设目标的事情，我们合起来导致每个人的M设目标会实现得更糟。这就像这样一种事例，如果每人所做的伤害不到什么人，我们合起来伤害到许多人。在这样的一些事例中，认为道德上重要的只是每个人所做的事情的后果是一个错误。我们必须同意，即便我们拒斥C而接受常识道德，这也是一个错误。

假设我们否定道德理论必须在集体性层次上是成功的。即便我们否定这一点，我们也必须承认，在我所讨论的那些事例中，常识道德犯了一个错误，因而应当予以修正。

40.一个更为简单的修正

常识道德有一个更为简单的修正，对此我尚未论及。这是该种道德的十足地行动者中性的形式，简称为N。按照此理论，我们每个人应当总是努力做总体上说来将最好地实现每一个人的M设目标的事情。我们的相对于行动者的道德目标成为共同的目标。

因为它局限于特定的一些事例，我所提议的修正R变得复杂起来。N是简单的，而且理论上更加具有吸引力。这提示着，如果我们已经移动到R的话，我们应当进一步移动到N。

可能看起来我本可延伸上面的论证，以便产生这个结论。考虑任何一个事例，其中某人要么会（1）促进他自己的M设目标，要么会（2）更为有效地促进他人的M设目标。所有这样的事例，放到一起，形成包括一切的事例（All-Inclusive Case）。它包括常识道德区别于N

的所有事例。如果我们应该在这个包括一切的事例中修正这个道德，我们就应该采纳N。

假设，在这一包括一切的事例中，每一个人做（1）而不是（2）。我们从而将使我们的M设目标总体上实现得更糟。如果我们全都做（2），那么它们总体上会实现得更好。但是这只是总体上说来或者对我们绝大多数人而言才会是真的。会有例外的。每个人的M设目标会实现得更好一些这一点不会为真。

回忆一下撒玛利亚人的两难。这是康德的《道德形而上学基础》中的第四个例子。我本可合理地希望没有人帮助深陷困境的陌生人这一点成为一个普遍的律法吗？对我们绝大多数人而言答案是否定的。但是有一些例外，诸如那些富甲一方或者有权有势的人，他们有贴身保镖和仆人侍候。这些人本可合理地希望没有人帮助深陷困境的陌生人。如果没有什么人帮助这样的陌生人对几乎每一个人而言都会更糟。但是它不会对每一个人都更糟。

一个类似的主张适用于我所说的包括一切的事例。如果我们全都成功地遵循了M的对行动者中立的形式的话，就绝大多数人的情况而言，他们自己的M设目标会实现得更好，这一点会是真的。但是这一点并非就每一个人而言都会是真的。按照我所给出的定义，M在此是直接集体性自败的这一点并非对每一个人都是真的。（换一种方式来表达这个要点。在那个定义中，“每一个人”意指“某个群体中的所有成员”。在包括一切的事例中，大多数人会处于这个群体中。但是会有一些圈外人。）

在运用康德的检验标准方面，我们或许可以坚持认为，富人和强者拉下了一个无知之幕。这可能被人们主张为道德推理的要求之一。但是我关于包括一切的事例却不能提出与之可比的主张。我的论证的目标不是对准那些前道德的人，把他们引向道德，而是对准那些相信常识道德的人们。既然这些人已经持有一种道德观点，我不能够类似地主张说，他们必须拉下一个无知之幕。这就是我没有为常识道德的

那个实足的对行动者中立的形式使用这个论证的原因。我只为限制更多的版本R提供了论证。R只应用于这样一些事例，就是，如果我们全都遵循M而不是R，我们将从而使每个人的M设目标实现得更糟。我们全都不仅以行动者中立的标准衡量做得更糟，而且以相对于行动者的标准衡量也做得更糟。这是一个关键性差异。在包括一切的事例中，我只能够主张，如果我们全都遵循M，我们将从而导致我们的M设目标总体上说来实现得更糟。主张这一点重要就是假定行动者中立的。这个主张可不能是行动者中立的一个论证的组成部分，因为这会将待证假定作为论据来进行辩论。

---

[1]参见Nagel（1），第127页。

[2]按照这些有关自我利益的理论，合乎理性与合理行动，无论它们的后果可能是什么，对各人而言都是好的。但它们只是那些对于我们而言是好的东西中的两个而已。因而总体上说来它们对于我们而言会是更糟。当假如我们合乎理性且合理行动的话，我们就会造成我们自己过多丧失其他对于我们而言是好的东西的时候，这就会为真。当这为真的时候，合乎理性与合理行动并不是S为我们设定的那个终极目标的组成部分。但是在那些囚犯的两难困境中这并不为真。在这些事例中，如果各人做S主张为合理的事情，那么这对于他将更好一些。假设我们接受这些理论之一。我们应当如何回答我的问题？如果我们两者都做A而不是S，以S的标准来衡量我们是在做得更好一些吗？我们是在造成各人的那些S设目标更好一些达到吗？答案是：在一个方面是肯定的，在另一个方面则是否定的。我们造成各自的生活在一个方面过得更好。我们使结果对于各人而言更好。但是我们造成各自的生活在另一个方面更糟了。我们是在非理性地行动。按照这些有关自利的理论，如果我们非理性地行动，本身就是糟的，无论后果可能是什么。
我们现在必须问，总体上说来，我们是否造成我们各自的生活过得更好。答案取决于该事例的具体细节。如果我们两者都做A而不是S，在多少程度上我们使结果对于各人而言更好了？如果我们使之在非常大的程度上更好，这将压倒我们是在非理性地行动这个糟糕的特征。如果我们只是使结果略微好一些，这将不会超过那个糟糕的特征。回忆一下那个原始的例子。如果我们两者都保持沉默，各人将以给自己添加两年徒刑的代价省去另一个人10年牢狱之灾。净后果是各自将被省去8年的牢狱之灾。这个后果对各人都更好。我们以非理性地行动为代价达到这个后果。如果我们以这个方式非理性地行动对于我们而言是糟糕的，那么这就像度过8年牢狱之苦一样糟糕吗？我们可能说"更糟"。按照这个观点，该事例不是一个真正的两难。下面这一点并不是真的：如果两者中各人而不是没有人做对于他自己而言将更好的事，那么对于我们两者而言将更糟。按照我们的观点，如果他认罪而不是保持沉默，那么对于各人而言将更好，因为他将减去自己两年的徒刑。但是，如果两者都坦白认罪而不是保持沉默，那么对于各人而言就不会更糟。如果两者都保持沉默，那么我们以非理性地行动为代价

减去各人8年徒刑。按照刚刚提及的那个观点，这对于各人而言将更糟。以这种方式非理性地行动对于各人而言比在监狱中度过8年要更糟。

在这个事例中，这个答案看起来可能是荒谬的。但是，假设我们以非理性地行动为代价减去各人8小时的牢狱之苦，或者8分钟的坐牢时间。那么主张该事例并不是一个真正的两难这一点就比较不怎么显得荒谬了。

[3]许多作者都已经顺着这些话加以论证。参见，例如Baier（1）和（3），Gauthier（3）和（4）。

[4]参见Schelling（2）。

[5]例如，这一点就Warnock，Mackie（2），Baier（1），Brandt（2），Harman，Gert，Toulmin，Rawls和许多别的人的观点而言是真的。

[6]（此注为1987年所加）在其刊载于1986年7月《伦理学》之中的那篇文章里，A·库夫里克论证说，常识道德并不包括我所称作M的东西的整体，因而并不是自败的。这可能如此。但是我的主张本来会是，假如常识道德的确包含了M的整体的话，它就会是自败的。更重要的是，除非这个道德包含我的修正R，否则，就将有许多协同难题，对此它令人不快地未能解决。（参见同一期中我的“回应”。）

第五章

结论

在本书第一编中我提出了当一个理论自败的时候表明什么的问题。让我们简略地回顾一下那些答案。

41.缩短常识道德和后果论之间的距离

在那些囚犯的两难困境中，自利论是直接集体性自败的。在这些事例中，如果我们全都追求自我利益的话，对我们所有的人而言将会更糟。如果相反我们全都合乎道德地行动的话，就会对我们所有人更好。一些学者论证说，因为这一点为真，所以甚至以利己的标准衡量道德论也优于自利论。

正如我在第四章中所表明的，这个论证不成立。在这些事例中，S在个体水平上是成功的。既然S是一个关于个人合理性的理论，那么它无需在集体水平上是成功的。

当这一论证被那些相信常识道德的人所提出的时候，它产生适得其反的结果。尽管它并未驳倒S，但是它的确驳倒了他们自己的理论的组成部分。像S一样的是，常识道德时常是直接集体性地自败的。与S不同的是，一个道德理论必须是集体性成功的。这些相信M的人们因而必须修正他们的信念，从M转移到R。他们的理想行动理论应当包括（R1），而且他们的实践行动理论应当包括（R2）和（R3）。

与M不同的是，R是后果论的，为我们大家设定了共同的道德目标。既然第四章表明，M信奉者必须移动到R，这就缩小了常识道德与后果论之间的分歧。

第一章也缩小了这一分歧。如果我们总是力求直接遵循C，做无论任何会使结果最好的事情，那么我们就是纯粹的行善者。C是间接地集体性自败的。如果我们都是纯粹行善者的话，那么结果会比倘若我们具有某些别的愿望和意向的话要糟糕。这个事实并未驳倒C；但

是它表明C必须包括理想动机理论和实践动机理论。C的理想动机理论必定主张，我们不应当都成为纯粹的行善者。C的实践动机理论必定主张，我们每个人应当力求具有以C的标准衡量的尽可能最好的愿望和意向的集合之一。（如果没有什么别的可能的集合这一点为真，即如果这个人具有这一别的集合，结果会更好的话，那么某人就具有这些最好的可能集合之一。）

对我们大多数人来说，最好的那些意向会在下述意义上与常识道德大致相呼应。我们应当时常强烈地倾向于做这种道德所要求的。

此点为真的方式有二。如果我们大多数人具有大多数幸福所系的那些强烈的愿望，那么将使结果更好。例如，如果我们爱我们的家庭和朋友，结果也将更好一些。那么我们将会强烈地倾向于给予这些人们的利益以特定种类的优先权，就像M主张我们所应该做的那样。如果我们绝大多数人具有做好自己工作的强烈愿望，也将使结果更好。那么我们将强烈地倾向于给予诸如我们的学生、病人、客户、顾客或者我们所代表的那些人的利益以某种优先权。以这些方式行动，我们再次会是在做M主张我们所应该做的事情。

出于不同的理由，我们应当强烈地倾向于不以某些特定的方式去行动。如果我们想要某个人死，我们很有可能将错误地认为，这个人的死亡会使结果更好。我们因而应当强烈地倾向于不要杀戮别人。类似的评述适用于欺骗或者强迫他人，适用于向威胁妥协，以及适用于常识道德所宣称为错的其他数种行动。

42.面向统一理论

因为C是一种行动者中立的理论，所以它是间接自败的，因而需要包括理想动机理论和实践动机理论，这些在上文所定义的意义上大致与M对应。因为M是一种相对于行动者的理论，它时常是直接自败的，因而需要得到修正，以使其理想行动理论和实践行动理论部分是

后果论的。C和M面对一些诘难，这些诘难只有通过扩大和修正这些理论，使它们相互更加接近，才能够予以答复。

这些事实自然地提示着一种吸引人的可能性。第一章和第四章中的论证所支持的那些结论可能组合形成或者联合形成一个更大的整体。我们或许能够发展出一个包括和结合起C和M两者的那些修正版本的理论，称之为统一理论。

这些主张就像西季威克、黑尔和其他一些人所提出的那些主张。[1]但是至少存在两个差异：

（1）这些作者中的绝大多数都力图将常识道德与功利主义（或缩写为U）结合起来，西季威克为U的快乐论版本作了论证；而黑尔则为愿望实现论版本作了论证。我一直在讨论那个更为广泛的理论——后果论。C可能诉诸数个有关什么使结果变得糟糕的原则。例如，C可能主张说，如果存在更多的不平等、欺骗和强迫，以及人们的权利得不到尊重和实现的话，就会更糟。如果C提出这些主张，与U相比，C就已经与常识道德更加接近了。

（2）在我提出的有关C是间接自败的这个主张方面，我只是遵循了西季威克和黑尔的观点。但是我在针对常识性道德的论证方面却没有遵循他们两者中的任何一个。这一论证并不像黑尔的论证那样诉诸一个有关道德性质的特定理论或者道德语言的逻辑。而且我的论证也并不像西季威克的论证那样诉诸我们的直觉。我主张的是，常识道德在许多事例中是直接集体性地自败的。除了常识道德所做的那些假定之外，这个主张不需要什么别的假定。当我得出在这些事例中这个道德必须予以修正的结论的时候，我的确假设一个道德理论必须在集体层次上是成功的。但是这个假定并非只由一种有关道德性质的理论所作出。它要么为有关这一主题的其他绝大多数不同理论所作出，要么为它们所蕴含。而且它会被那些相信常识道德的某个版本的人们所接受。

第一章和第四章中的论证都指向一个统一理论。但是要以一种令人信服的方式展开这个理论需要的篇幅至少会是一本书。彼书非此书。我将只添加某些简短的评述。

43.有待完成的工作

根据C，我们应当时常强烈地倾向于以M所要求的那些方式去行动。如果我们是C信奉者，而且我们既具有又按照这些意向行动，那么M信奉者不能够诘难我们所做的。但是他们能够诘难我们的信念。由于我们的意向是要以那些时常使结果更糟的方式行动，我们会时常相信，就像M信奉者认为我们应当做的那样，当我们按照这些意向行动的时候，我们会是在错误地行动。在发展出一个统一理论方面，我们最大的任务会是调和这些相互冲突的信念。

除了主张我们应当具有这些意向之外，C还可能主张，我们的政策应当是按照它们来行动。C可能主张，即便这些并非是那些要做无论什么使结果最好的事情的意向，这也应当是我们的政策。遵循这一政策会使结果最好这一点在第一章中所表述的那些方式方面可能是真的。其次谨记C的反应理论。这一理论主张，我们应该感到懊悔，应该责难别人，如果当这样做会使结果最好时。如果我们遵循刚刚所述的政策，我们时常将无法做到会使结果最好的事情。C从而将主张说，在某种意义上我们是在错误地行动。但是C也可能主张说，因为我们是在遵循这个政策，我们不应当受到责难，或者感到懊悔。C或许蕴含着，只有当我们并未遵循这个政策的时候，我们才应当受到责难，而且应当感到懊悔。这或许正是会使结果最好的责难和懊悔的模式。如果C提出这些主张的话，这会减弱C和M之间的冲突。尽管这些理论在有关行动的对错方面仍然会存在歧见，但是在哪些行动是值得责难的和应当在我们之中引起懊悔方面它们的歧见则要小得多。我们会更加接近那个统一理论。

最后这些主张有些过于简单化。在发展那个统一理论方面，我们会既需要考虑不同种类的行动和政策，又需要考虑这些会如何与诸如我们的情绪、需要和能力等事情相联系。许多问题需要得到答案。要成为令人信服的理论，统一理论必须作出许多区别，并且提出许多不同的主张。会有大量的工作有待人们去做，但在此我暂不作任何尝试。

既然统一理论会包括C的一个版本，它可能遭到诘难，会说它要求太高。但是这个诘难也可能由C的反应理论给予部分回答。

回到那个有关富裕国家的人们应当给予那些穷人多少的问题。既然其他人事实上将给出甚少，那么C主张，每一个富人应当给出他的收入中的几乎所有部分。如果富人所给较少一些，那么他们就是在错误地行动。但是如果每个富人因为未能给出他的几乎全部收入而受到责难，那么责难就会不再有效。责难和懊悔的最好模式是那个会使富人给出得最多的模式。既然如此，那么C或许蕴含着，只有当富人未能给出他们收入中小得多的一个部分，诸如十分之一的时候，富人才应当受到责难，才应当感到懊悔。

与常识道德相比，C在另一些方面则要求得更为多得多。于是C有时会主张，我们中的一个人应当牺牲他自己的生命，以便他能够拯救一些陌生人。但是未能拯救这些陌生人这一点即便以C的标准衡量也不会是值得责难的。既然统一理论包括C，那么它会比现在这样的常识道德要求更高一些。但是，如果它提出有关责难和懊悔的这些主张，它的要求则可能并非是不合理的或不切实际的。

44.另一种可能性

很多人是道德怀疑论者：认为没有什么道德理论能够是真的，或者是最好的理论。如果我们继续具有深刻的歧见的话，拒斥怀疑主义可能是相当艰难的。我们的那些深刻分歧之一是后果论者们与那些相信常识道德的人们之间的一个分歧。

第一章和第四章中的那些论证减弱了这个分歧。如果我们能够发展出那个统一理论的话，这个分歧可能会被消除。我们可能发现，用穆勒的话来说，我们的对手“正在另一面山坡爬山”。

因为我们的道德信念不再有分歧，所以我们或许也会改变我们有关这些信念的地位的观点。道德怀疑论可能从根本上遭到削弱。

---

[1]参见Sidgwick（1）第三编和第四编，（以及Schneewind著作中的述评），Hare（1）、Sumner（1）和Sumner（2）第五章。


第二编 合理性与时间
上海译文出版社

第六章

对自利论的最好诘难

45.当前目标论

第一编中的论证并未驳倒自利论。现在我将提出其他一些针对这一理论的论证。这些论证中的某一些诉诸的是道德。但是主要的挑战来自有关合理性的一个不同理论。这就是当前目标论（缩写为P）。

我描述过P的三个版本。其一是工具论（缩写为IP）。这一理论主张

IP：我们每个人最有理由做的是最好地实现其当前愿望的任何事。

我延伸了"愿望"一词，使之涵盖意图、设想以及其他目标。

要应用工具论，我们必须既能够区分不同的愿望，又能够决定某个人真正具有的是哪一些愿望。两者都可能是困难的。但是我在此将忽略这些困难。我所需要假定的一切只是我们有时能够决定，总体上说来什么会是最好地实现某人的当前愿望。

在决定这一点方面，我们应当忽略那些派生的愿望。这类愿望是用以实现其他愿望的纯粹手段。假设我要到某个图书馆，只是为了我能够碰到某个漂亮的图书馆馆员。如果你把我引见给这个图书馆馆员，我就没有什么未实现的愿望了。你没有实现我加入这个图书馆的愿望这一点是不相关的。我将用"愿望"指"非派生的"愿望。

在决定什么会最好地实现这些愿望方面，我们应当给予那些更为强烈的愿望以更大的权重。某个人的最强烈的愿望可能被数个其他愿望在权重上所超过。假设，如果我做了X，就不会实现我最为强烈的当前愿望，但是却会实现我的所有其他当前愿望。尽管X不会实现我的那个最强烈的愿望，但是我可能断定，就我可能做出的那些行动而

言，X总体上说来会最好地实现我的诸愿望。如果我对这一点作出决定，那么把一切考虑在内，X就可能成为我最想要做的事。

在其处理愿望之间的冲突方面，以及在其面临风险和不确定因素作出决断方面，工具论可能呈现出一些微妙的形式。但是正如其名称所蕴含的，它完全关注手段的选择。它并不批判行动者的目的——他所抱有的愿望是什么。正如休谟的众所周知的话："人如果宁愿毁灭全世界而不肯伤害自己一个指头，那并不是违反理性。如果为了防止一个印第安人或与我完全陌生的人的些小不快，我宁愿毁灭自己，那也不是违反理性。我如果选择我所认为较小的福利而舍去较大的福利……那也同样不是违反理性。"[1]

拒绝对愿望加以批判，这并不是当前目标论的一个本质部分。休谟甚至提出，"一种情感……当依据某种虚妄的判断时……可以说是'不合理的'"。这一提法在审慎论（缩写为DP）中得到发展。这一理论主张

DP：我们每个人最有理由做的，不是能够最好地达到他实际上想要的东西的事，而是，在行动的时候如果他已经经历"理想的审慎"过程——如果他知道相关的事实，思考清楚，而且不受歪曲的影响——能够最好地达到他本会要的东西的事。

相关的事实是以此为真的那些事实，那就是，如果这个人知道这个事实的话，他的愿望会变化。最后这个主张需要以我们在此可以忽略的一些方式予以提炼。[2]

P的第三个版本是批判性的当前目标论（缩写为CP）。这一理论主张，某些愿望是内在地非理性的，并不为行动提供好的理由。CP还可能主张，某些愿望是合理地要求的。按照这第二个主张，如果某人并不具有这些愿望的话，他就会是非理性的。

我们在此必须区分两类理由：解释性的和好的。如果某人以一种特定的方式去行动，我们可能知道他的理由是什么。通过描述这个理由，我们解释这个人为什么像他所做的那样去行动。但是我们可能认

为这个理由是一个非常糟糕的理由。我将用“理由”意指“好的理由”。按照这个用法，我们会主张说，这个人没有像他那样去行动的理由。

按照审慎论，任何一个愿望都提供一个行动理由，如果它经受住了理想的审慎过程的话。假设，在知道事实并且思虑清楚的情况下，人们宁愿毁灭全世界而不肯伤害自己一个指头。按照审慎论，如果我有一部末日机器，而且按照这样的一种取舍而行动，这样做会是合理的。我们可能拒斥这个主张。我们可能认为，这个取舍并不为行动提供一个理由。而且有关其他许多可能的愿望，我们可能抱有同样想法。按照审慎论，没有什么愿望是内在地非理性的。如果我们相信有这样的一些愿望，我们就应当拒斥这个理论。

一位审慎论者或许回应说，这样的非理性愿望经受不住理想的审慎过程。凭借下定义的方法，能够使此一回应为真，但微不足道。该理论家可能主张，任何具有这些愿望的人不能够“思虑清楚”。但是在后果上这等于认可了该诘难。在界定什么算作“思虑清楚”的时候，该理论家会不得不指称正在谈论中的那些愿望。他会不得不作出哪些愿望属于内在地非理性的判断。

审慎论者可能把其回应代之以一个事实主张。他可能同意，他所说的“思虑清楚”在逻辑上并不排除具有那些我们所认为的内在地非理性的愿望，但他可能坚持认为，他的理论是适当的，因为那些思虑清楚并且知道那些事实的人们不会具有这样的愿望。

这一点是否为真难以预测。而且，即便此点为真，我们的诘难也不会得到完全答复。如果某些特定种类的愿望是内在地非理性的，任何一个完备的有关合理性的理论就都应该主张这一点。我们不应当忽略那个是否存在这样的愿望的问题，只是因为我们希望，如果我们思虑清楚，我们将永远不会具有它们。如果我们认为能够有这样的愿望，我们就应当从当前目标论的审慎论的版本转移到其批判论的版本。

这个理论令人奇怪地受到忽视。被人们广泛相信的工具论的和审慎论的版本，有两个主张：（1）每一个人最有理由做的乃是会最好地实现这样一些愿望的事，即在行动的时候，如果他知道事实而且思虑清楚的话他所具有或者本会具有的那些愿望。（2）愿望不可能是内在地非理性的或者被合理地所要求的。这些是十分不同的主张。我们会拒斥（2）而采纳（1）的一个加以限定了的版本。那么我们会接受P的那个批判论的版本。这一版本主张

CP：某些愿望是内在地非理性的。而且一组愿望可能是非理性的，即便该组中的那些愿望并非是非理性的。例如，宁取X而舍Y、宁取Y而舍Z，以及宁舍Z而取X，是非理性的。一组愿望也可能是非理性的，因为它未能包含那些合理地所要求的愿望。假设我知道那些事实并且是思虑清楚的。如果我的那组愿望不是非理性的，我最有理由做的乃是会最好地实现我当前的那些并非非理性的愿望的事。这个主张适用于任何时间的任何人。

"非理性的"这一指控处于一个批判域的一端。它就像"邪恶的"这一指控一样。我们可能主张，某个行动尽管坏但不至于邪恶，仍然易受道德的批判。同样我们可能主张，某个愿望尽管不至于遭受"非理性的"这一极端的指控，但是易受理性的批判。为节省文字起见，我将扩展"非理性的"一词的普通用法。我将用该词意指"易受理性批判的"。这将允许"并非非理性的"意指"并非易受理性批判的"。

在其有关那些并非非理性的愿望的主张中，CP无需只是诉诸这些愿望的强度。例如，它可能并不给予某人但愿他自己并不具有的那些愿望以什么权重。而且CP无需只是诉诸甚至在最为广泛的意义上我们才能够称之为愿望的东西。它还可以诉诸行动者的价值、理想或他的道德信念。所有这一切都能够为行动提供理由。但是CP主张这些理由中的一些不是好的理由。例如，它可能主张，即便对那些相信礼仪或者某种荣誉规范的人们而言，也没有什么理由遵从某些毫无意义的规矩或者进行荣誉所要求的决斗。

我已经描述了当前目标论的三个版本。尽管这一描述只是一个梗概，但在此足够了。下面的相当篇幅将关于这些不同版本之间的共同

之处。部分出于这一原因，我将只讨论审慎论和工具论相吻合的那些事例。这些是某人在其中知道所有的相关事实而且思虑清楚的那些事例。我还将假定，会最好地实现这个人的当前愿望的东西与这个人在一切都考虑到的情况下最想要的东西是一样的。而且我将时常假定这个人的愿望与他的道德信念或者别的什么价值和理想并不冲突。通过作出这些假定，我得以避免考虑几个重要问题。这些问题必须得到任何一个完备的有关合理性的理论的回答。但是它们与我在本书第二编中的主要目标并不相关。这个目标在于表明我们应当拒斥自利论。

既然这是我的主要目标，那么第二编对于那些业已拒斥这个理论的人们而言可能是枯燥的。但是我还将讨论有关合理性与时间的某些令人困惑不解的问题。而且我将支持CP，主张某些愿望是内在地非理性的，而且其他一些愿望可能是合理地所要求的。既然这些主张具有争议性，那么在回到自利论之前我将给予它们某种辩护。

46.愿望能够是内在地非理性的抑或是合理地要求的吗?

人们为什么认为，如果一个愿望并非基于某个虚假的信念，那么它就不可能是非理性的？我们首先可以回忆一下著名的休谟为什么持有这一观点。休谟把推理当作只是与信念和真假有关的。一个愿望不能够是虚假的。按照休谟的观点，一个愿望只有在它包括理论性的非理性的时候才能够“被称作不合理的”。

推理并非只是关乎信念。除了据以信仰的理由之外，还有据以行动的理由。除了理论上的合理性之外，还有实践上的合理性。于是就有了一个与之不同的、更简单的方式，其中一个愿望可能是非理性的。它可能是一个并不为行动提供理由的愿望。

休谟的一些追随者拒绝称愿望是“非理性的”。如果这些人同意这一点，即某些愿望并不为行动提供好理由，那么这个差异就会是微不足道的。其次谨记我用“非理性的”意指“易受理性批判的”。这是一个程度问题。如果一个愿望并不完全是非理性的，那么它可能为行动提

供某种理由。如果两个愿望之一是更易受理性批判的，那么它所提供的理由则更弱一些。

如果一个愿望是完全非理性的，那么它并不直接地为行动提供任何理由。但是一些愿望，尽管是非理性的，却的确间接地提供这样的一些理由。如果我患有幽闭恐惧症，我可能具有强烈地不愿处于某个封闭空间内的愿望。这个愿望就像一种恐惧。既然恐惧涉及一个有关所恐惧的对象是危险的这样的信念，那么当这个信念显然是虚假的时候的恐惧是非理性的。假设，当我具有逃离某个封闭空间的强烈愿望时，我知道我并不处于什么危险之中。既然我的愿望就像一种恐惧，那么我可能判定它是非理性的。但是这个愿望间接地为行动提供了一个理由。它使我强烈地不喜欢处于这个封闭空间之中，而且我具有力求逃离我所强烈地不喜欢的东西的一个理由。

当一个愿望直接为行动提供一个理由的时候，该理由很少就是该愿望。这样的情况很少为真，那就是，当某人以某种方式行动的时候，他的理由只是他想要这么做。在绝大多数事例中，某人的行动理由是他所想要的东西的特征之一，或者是解释并证明其愿望正当性的事实之一。假设我帮助一个需要帮助的人。我帮助这个人的理由并不是我想要这么做，而是他需要帮助，或者我承诺过要提供帮助，或者诸如此类。同样，我阅读一本书的理由并不是我想要这么做，而是这本书充满机智或风趣，或者它解释了为什么宇宙存在，抑或它在其他某个方面值得一读。在这两个事例中，我的理由并不是我的愿望，而是我所在做的之所以值得一做的那个方面，或者我的目标之所以是可欲的（desirable）——值得欲求的——那个方面。[3]

如果一个理由很少就是一个愿望，那么这可能看起来从根本上削弱了当前目标论。可能看起来这表明我们最有理由做的不能够依赖于我们所欲求的。但是情况并非如此。即便一个理由不是一个愿望，但是它可能依赖于一个愿望。假设我阅读某本书的理由是它解释了第一

次世界大战的那些原因。如果我没有想知道这些原因是什么的愿望，我就不会有理由去阅读这本书。

我在上文主张，根据CP，某些愿望可能是合理地所要求的。回到我帮助某个人的理由是他需要帮助的那个例子。这个理由依赖一个愿望吗？即便我并不在乎这个人的需要，我会有理由帮助这个人吗？更宽泛一些说，即便我并不在乎道德，我会有理由道德地行动吗？

这些都是具有争议性的问题。一些学者对两者的回应都是否定的。根据这些学者，这两个理由本质上都依赖我们的愿望或者我们所在乎的东西。[4]

另一些学者主张，我的确具有理由道德地行动或者帮助某个需要帮助的人，即便我并没有以这些方式行动的什么愿望。这一主张与P的工具论的版本相冲突，而且可能与审慎论的版本相冲突。但是它不一定与批判论的版本相冲突。如果我们接受CP，我们可能主张

（CP1）我们每个人都合理地被要求既要在乎道德又要在乎他人的需要。有鉴于此，我们具有道德地行动的理由，即便我们没有这么做的什么愿望。我们是否有理由以某个特定方式行动通常取决于我们是否具有某些特定的愿望。但就被合理要求的那些愿望的事例而言则并非如此。

这可能会被接受CP的某个人所主张。是否应当如此主张的问题在我对CP的描述中仍然悬而未决。既然这个问题具有争议性，最好还是不作定论。两种回答中的任何一个答案都可能会被某个接受CP的人所给出。如果一个理论对每一个争议性问题都悬而未决，它就不会值得讨论。但是，就像我们将看到的，CP的情况并非如此。

CP的另一个与众不同的主张是，某些愿望或者某些愿望集合是内在地非理性的。我在上文中写道，在绝大多数情况下，我的行动理由并不是我的愿望之一，而是我所欲求的之所以值得欲求的那个方面。这自然提示着某个愿望如何会是内在地非理性的。我们能够主张："欲

求某个没有什么方面值得欲求的东西是非理性的。欲求某个不值得欲求——要加以避免——的东西更是非理性的。”

或许有人说，性受虐狂具有这样的愿望。但实际的性受虐狂是一个复杂的现象，需要冗长的讨论。我们可以设想一个比较简单一些的事例，其中某人只是要求在未来的某个时刻遭受巨大的痛苦。假设这个人与受虐狂们不同的是，他明白：他不会在哪一方面享受这一疼痛，不会发现它会降低他的罪恶感，也不会在任何别的方面从这一疼痛中受益。这个人只是想要拥有他在那时将强烈地不喜欢而且非常强烈地想不要具有的那些感受。我们绝大多数人会认为这个愿望是非理性的。

实际上存在着一些人们具有非理性愿望的事例吗？一个例子可能是这样一种愿望，那就是，当一些人身处悬崖边缘的时候所具有的那种要跳下去的愿望。这一奇异的冲动甚至那些丝毫都不想死的人们都会有过。既然这些人想继续活着，那么对他们而言按照他们要跳下去的愿望行动可能就是非理性的。但是这并不表明这个愿望是非理性的。跳下去的愿望并不是一个去死的愿望。就某些人的情形而言，它是一种在空中翱翔的愿望。这是某种值得欲求的东西；我们能够合理地羡慕飞鸟。而就其他一些人的情形而言，尽管他们想要跳下去，但是他们丝毫不具有在空中翱翔的渴望。就这些人们的情形而言，他们的要跳下去的愿望可能是非理性的。

接下来考虑一下这个愿望的另一种形式。据说，在他们迷狂出神的巅峰，某些日本夫妇会从悬崖奋力跃下，因为他们的确想死。他们想死，是因为他们处于迷狂出神的巅峰。这能够是一个自杀的理由吗？

一些人会说：“不是。死亡因为它将终结一个人的极端痛苦而可能值得欲求。但是死亡不能因为将终止一个人的迷狂出神而值得欲求。”

对该事例的这一描述是错误的。这些夫妇并不是因为这将终结他们的迷狂出神而想死。他们想要他们的生命在最高点或者最好点结

束。这不是我们大多数人所想要的。但是，尽管这个愿望非同寻常，但它并不是清清楚楚地非理性的。迷狂出神并不会持续良久，而是会下降和消退。如果某对夫妇处于迷狂出神之中，他们能够令人信服地把它的自然消退当作某种非常不值得欲求的东西或值得避免的东西。通过中止他们的迷狂出神，他们的死亡会确保这一迷狂出神不会衰退。对这样的夫妇而言，死亡可能是值得欲求的。

还有其他一些方面，那些看起来疯狂的愿望可能并不是非理性的。例如，这些愿望的对象可能在美学上具有吸引力。考虑一下那些奇思妙想。内格尔写道："一个人也许根本没有什么理由设想这样一种愿望，就是月亮上会有西芹，他会想在下一次火箭升空时偷偷带点放进去；一个人或许只是喜欢这一想法。"[5]这个愿望并不是非理性的。它是一个极佳的奇思妙想。（作为对照，大海里长西芹作为奇思妙想则大为逊色。）

对某种在任何方面都不值得欲求的东西，或者值得避免的东西抱有愿望是非理性的。尽管我们能够轻易地想象出这样的一些愿望，但是以这种方式非理性的愿望实际上可能少之又少。而且有一大类的愿望不可能是非理性的。这些是在纯粹的身体痛苦或者快乐之中所涉及的那些愿望。我喜欢冷水浴。他痛恨冷水浴。没有哪一个是非理性的。如果我因为其味道而想要吃某种东西，这个愿望不可能是非理性的。即便我所喜欢的会引起他人的恶心，它也不是非理性的。其次考虑一下我们发觉并不愉快的那些经验。许多人具有不想听到粉笔在黑板上写字发出吱吱嘎嘎声音的强烈愿望。这个愿望是奇怪的，因为这些人并不介意听到那些在音质和音高方面与之相似的吱吱嘎嘎的声音。但是这个愿望并不是非理性的。类似的主张适用于我们发觉充满痛苦的东西。

现在转向我们有关不同的可能痛苦和快乐的愿望。正是在这个第二层次上"非理性的"这一指控才可能是最令人信服的。某人只是因为发觉一个经验比另一个经验更痛苦并不就是非理性的。但是，如果当

他不得不经受这两个经验之一的时候，他宁取那个将更为痛苦的一个，他则可能是非理性的。这个人也许能够为这一取舍辩护。他可能认为，作为某种忏悔形式，他应该遭受更大的痛苦。或者他可能想使他自己变得更顽强一些，以便能够更好地忍受以后的痛苦。或者他可能认为，通过审慎地选择现在经受两个痛苦中更甚一些的那一个痛苦，而且坚持这一选择，他将把自己的意志磨炼得更加坚强。抑或他可能认为，较大的受难将启迪人的智慧。在这些方面和其他一些方面，某人对遭受两个痛苦中更甚的那一个痛苦的愿望可能并不是非理性的。

考虑一下下面这个想象出来的事例。某个快乐主义者很在乎他未来经验的品质。除了一个例外之外，他同等地在乎他的未来的所有部分。这个例外就是，未来一周二一漠不关心。每一个周二他通常对他身上发生什么都很在乎。但是他对某个未来的周二所可能有的痛苦或者快乐从不在乎。于是他在接下来的周二会选择一个充满痛苦的手术，而不是在接下来的周三选择一个痛苦少得多的手术。这个选择不会是任何虚假信念的结果。这个人知道，如果是在周二该手术将更加痛苦。他也没有个人同一性的虚假的信念。他同意，周二将遭受痛苦的就是他。他也不是对时间具有虚假信念。他知道周二只是传统历法中的一个部分，其取自一个虚假的宗教名称也具有随意性。他也没有什么别的信念或许有助于证明其对未来周二的痛苦漠不关心的正当性。这一漠不关心是一个赤裸裸的事实。当他在筹划自己的未来的时候，他总是宁取周二巨大的受难前景而不是别的任何一天最为轻微的痛苦这一点完全属实。

此人的这种在乎模式是非理性的。他为什么宁取周二的极痛而舍任何一天的微痛？只是因为极痛将是在周二。这不是什么理由。如果某人必须在周二的极痛或周三的微痛之间作出选择，极痛将是在周二这个事实不是什么宁可选取它的理由。没有什么理由而宁选两个痛苦中更糟的那一个是非理性的。

这点可能遭到诘难说，因为此人的取舍是纯粹想象性的，而且是如此稀奇古怪，我们不可能有成效地讨论它是否是非理性的。因而我将比较其他两种对待时间的态度。其中一个是极为常见的：更在乎较近的未来。称之为近期偏向。某个具有这一偏向的人可能有意识地选择数星期之后较重一些的痛苦，而不是今天下午较轻一些的痛苦。这种选择时有发生。如果两个痛苦中更糟的那一个会在更远的未来的话，这能够是选择这个痛苦的理由吗？近期偏向是非理性的吗？许多作者主张它是非理性的。

下面考虑一下某个具有来年偏向的人。此人对来年一年中的未来关心的程度相等，而对他的未来的其他部分的关心却只相当于对来年关心程度的一半。再次，这个所设想的人没有什么有关时间、人的同一性或其他任何方面的虚假信念。他知道从现在算来一年以后生活的那个人正是他，而且在随后的那些年份中的痛苦程度将是与之一样的。

没有什么人具有此人的这种关心模式。但是它非常接近于那个常见的模式：近期偏向。差异在于这一常见偏向与它所珍爱的特征成正比。那些具有这一偏向的人们对更近的未来中的事情更加关心。我所设想的人所具有的近期偏向并不呈现为成比例的形式，而是呈现为一种粗陋的两个步骤形式。此人的偏向划出一道随意线。他对接下来的12个月的关心程度是同等的，而对任何随后月份的关心程度却只相当于前者的一半。于是他会有意识地选择距今13个月的3周痛苦而不是距今11个月的2周痛苦。当问及他为什么偏选更长时间的煎熬的时候，他说，“因为那是未来一年多以后的事”。这就像这样一个主张，就是，“因为它在更远的未来。”但是它是更易受到理性批判的。如果某个痛苦是在更远的未来，那么认为这是现在较不关心这个痛苦的理由恐怕是可以得到辩护的。但是难以相信以下情况可能是合理的，那就是，一方面对接下来12个月之内的所有痛苦同等关心，另一方面对所有随后的痛苦的关心程度却只有对前者的一半。如果某个痛苦将在

53个星期之后而不是52个星期之后被感受到，“这如何能够成为现在对它的关心只相当于对作为第一步的那12个月内的疼痛的关心程度的一半的理由？”

再举一对类似的例子。许多人对发生在他们邻居们身上或者他们自己社区的人们身上的事情更加关心一些。不会有多少人主张说这一关心模式是非理性的。一些人主张这是道德上所要求的。但是考虑一下一个其关心模式是1英里内利他主义的人。此人对距他家1英里之内的所有人的福利都非常关心，但是他对那些远于这个距离的人们则不怎么关心。如果他得悉1英里之内的人们遭受火灾或者水灾的时候，他将慷慨地为基金会捐款来帮助这些人们。但是如果这样的一些人处于1.25英里之外，他将不予以帮助。这并不是对此人的慈善之举选定限度的一个政策。它是此人对他人的受难的关心程度方面的真实差异所造成的结果。

此人的关心模式粗略地类似于那一常见模式：更关心我们自己社区的成员。但是他的关心划下另一条随意线。如果某一个人对他人没有什么关心，此点尽管可悲可叹，但是可能并不是非理性的。无论某人对发生在每一个人身上的事情同等关心，还是对发生在他自己的社区中的成员身上的事情更加关心，都没有哪一个是非理性的。但是如果某人对发生在距家不足1英里内的人们身上的事情极为关心，而对那些更远一些的人们的关心程度则少得多，那么这个关心模式就是非理性的。两个受难的陌生人，一个处于1英里之内而另一个处于1英里之外，这能够有什么区别？两个人中一个处于一英里之外这一点并不是什么应该受到较少关心的理由。

休谟的追随者们主张，如果一个关心的愿望或者一个关心的模式并不涉及理论上的非理性，那么它就不能够是易受理性批判的。我已经否定了这个主张。当其应用于身体上的痛苦和快乐所牵涉的那些愿望的时候的确如此。但是就某些一阶愿望而言可能并非如此。这些愿望中的一些可能是非理性的。在悬崖边想要跳下去的愿望可能就是一

个例子。如果这不是在空气中翱翔的愿望或抑制一个人的迷狂出神衰退的愿望，那么它就可能是非理性的。

当我们转向有关可能的痛苦和快乐的二阶愿望的时候，能够发现一些极好的例子。如果这样的愿望在同等好的快乐或同等糟的痛苦之间以一种随意的方式区别对待，那么它们就是非理性的。因未来的痛苦要么在一个周二感觉到要么在一年多以后的未来感觉到之故，而对这些未来的痛苦关心得少一些，则是非理性的。而且因他人处于1英里以外之故而对这些人的受难关心得少一些也是非理性的。在这些事例中，所关心的程度之所以较小，不是因为所关心的对象有某种内在差异。关心得少一些是因为一种纯属位置性的属性，而且划出了一条随意线。这些关心模式是非理性的模式再清楚不过了。这些关心模式是想象性的。但它们是那些常见模式的更为粗略的版本。如果未来的痛苦是在更远的未来，那么许多人对未来的痛苦关心得更少一些。此点时常被认为是非理性的。我将在第八章中讨论这一主张。[6]

47.三个彼此竞争的理论

现在回到自利论（S）。S是如何受到当前目标论或P的挑战的呢？S主张

（S2）我们每个人最有理由去做的是任何对他而言最好的事，而且
（S3）对任何一个人而言做他认为对己更糟的事是非理性的。

对P的一个论证可能迫使S退到更弱的主张。对S的威胁的严峻程度取决于两件事：那些论证的强度，以及如果它们成功的话将迫使S倒退多远。

最为雄心勃勃的威胁会是这样一个论证，就是，它表明每当S与P冲突的时候我们没有什么理由遵循S。如果这使我们在行动之时既知道事实又思虑清楚的情况下最想要或者最珍视的东西落空的话，我们

没有什么理由以符合我们自己利益的方式行动。对S而言这是一个彻底的失败。

我相信，我的那些论证证明了这一结论的一个版本的正当性。但是它们所表明的东西可能要少一些。它们可能只是表明，当S和P相冲突的时候遵循其中的哪一个都会是合理的。尽管这是一个较弱的结论，但是我将予以主张，对S而言它几乎是损伤性的。

我将提出几个论证。这些论证能够以一个战略性的隐喻予以引入。正如我们将看到的，自利论居于道德与当前目标论之间。它因而面临一个经典危险：两条战线作战。尽管它在只遭受一个方向的打击时能够得以幸存，但是在两面夹击的情况下它则可能不能幸存。我相信情况是这样的。许多作者论证说，道德提供最好的或者最强的行动理由。在拒斥这些论证方面，一位自利论者作出一些假定，而一位当前目标论者可以用以转而针对他。而且他对该当前目标论者的那些回应，如果它们是有效的，则从根本上削弱他对道德的诘难。

如果我们相信一个理论我们据以行动是合理的，那么我们则说这个理论在我们看来是幸存下来的。一个理论如果是惟一的幸存者，那么该理论则会取得胜利。那么我将相信，不按照这个理论行动则是非理性的。如果一个理论并未取得胜利，不得不承认那些没有被击败的对手的话，它则必须限定它的主张。就三个理论来说，可能会有8种结果。幸存者可能会是

（1）	道德	自利论	当前目标论	三头政治
（2）		自利论	当前目标论	
（3）	道德		当前目标论	两头政治
（4）	道德	自利论		
（5）	道德			
（6）		自利论		寡头政治
（7）			当前目标论	
（8）				无头政治

按照上述的那个较弱的结论，自利论并不能击败当前目标论。如果S幸存下来，P也幸存下来。这一点淘汰掉（4）和（6）。S只在（1）和（2）中幸存下来。但这些却被我所说的那个更强一些的结论所淘汰。而且我将主张，如果S只在（1）和（2）中幸存下来，那么对S而言这无异于一种失败。（本书对于（3）、（5）、（7）或者更加惨淡的（8）都着墨不多。）

要完成我的第一个论证，我们必须避免某些错误。如果我们像许多学者那样忘记自利论有两个对手——它既受到道德论的挑战，又受到当前目标论的挑战，那么这些就更加难以避免了。如果我们把自利论只与其一个对手相比较的话，我们就可能无法注意到一个在什么时候剽窃了另一个的论证。

48.心理自我主义

一个错误是假定自利论和当前目标论总是相契合的。没有人在P的工具论的版本方面作出这一假定。人们实际上想要的与其利益严重背离的事例实在太多了。但是人们普遍假定，P的审慎论的版本与S是相契合的。人们普遍认为，如果每个人真正知道事实而且思虑清楚的话，那么他最想要的会是去做对他而言最好的任何事，或者会是最好地促进其长远的自我利益的任何事。这一假定被称作心理自我主义。如果我们作出这个假定，那么把P只是当作S的一个部分可能是自然而然的事情。尽管是自然而然的，但这却是另一个错误。即便它们总是相契合，这两个理论也仍然会有区别。而且如果我们把P淹没在S之中，我们就可能无法按照S自己的功过来对S作出判断。它的一些可取性可能是从P剽窃而来的。

心理自我主义能借助定义而使之为真。一些学者主张说，如果某人想做的是他所知道的对他自己而言更糟的事情，那么他不可能是在“清楚地”思虑，而且必定是受到某种“歪曲的影响”。当该主张以这一

方式成为真的时候，它就变得微不足道了。P失去其独立性，而且按照定义来说与S相契合。这个版本的P并不值得讨论。

心理自我主义按照定义成为真的方式还有另外两种。一些学者主张，（1）对某人而言最好的东西按照定义是此人在既知道事实又思虑清楚的时刻最想要的东西。另一些学者主张，（2）如果某个行动会最好地实现某人的当前愿望，那么这个行动按照定义是使这个人的功利达到了最大化。再一次，当心理自我主义按照定义成为真的时候，它就变得微不足道了。按照这两个定义，失去其独立性的是S。（1）使S与那个P的审慎论的版本相契合；（2）使S与那个工具论的版本相契合。S的这两个版本并不值得讨论。显然，定义（2）所涉及的并不是按照任何一个令人信服的自利论来说什么符合我们自己的长远自我利益。正如我现在将要论证的，就定义（1）的情况而言同样如此。[7]

我们绝大多数人在绝大多数时间里都强烈地想要采取符合自己利益的行动。但是在许多情况下，这并不是某个人最强烈的愿望，而且即便是的话，它也被其他一些愿望在权重上所超过。在很多场合，即便就某个知道相关事实而且思虑清楚的人来说这一点也为真。例如，在当前目标论支持与自利论相冲突的道德的时候就是如此。某人最想要的可能是尽职尽责，即便他知道这将有悖于他的利益。（谨记，为简单起见，我们是在考虑某人最想要的东西与最好地实现其当前愿望的东西相同的那些事例。）还有许多并不牵扯到道德的事例，甚至在理想的审慎之后某人最想要的东西并不会与最有效地促进这个人自己的长远自我利益的东西相契合。这些事例中有许多是有争论的或者是晦涩不清的。但是正如我们将看到的，其中还有别的许多是清楚的。

这些事例的共通之处部分地取决于我们有关自利的理论。正如我在附录C中所主张的，这些事例按照快乐论来说共通之处要大一些，按照成功论则要小一些。按照非限定的愿望实现论，这些事例可能是最罕见的。按照这一理论，我们的任何一个愿望的实现都直接算作是

符合我们的利益的。最符合我的利益的东西是最好地实现或者能够使我实现我一生的所有愿望的东西。这与最好地实现我倘若知道事实并且思虑清楚的情况下的当前的愿望总是相同的吗？可能有许多人，对他们而言这两者通常是相契合的。但另有许多人，对他们而言两者时常是相冲突的。在这些人的生活中。即便S假定非限定的愿望实现论，S也经常与P相冲突。S和P之所以冲突，是因为这些人们的那些最强烈的愿望在他们的一生中并不是同一些愿望。

有一个复杂之处。这点涉及我在第三节中所讨论的人：对他们来说S是间接自败的。在其有关这些人的那些主张中，S与P发生冲突的方式较不直接。但是我们可以忽略这一复杂之处。我们可以在S并非间接自败的那些事例中讨论S。集中于这些事例对S不可能是不公正的。而且此处的那些重要问题呈现出更加清晰的形式。

心理自我主义不能经受细致的讨论而幸存下来。按照所有关于自利的令人信服的理论，S和P经常冲突。行动的时候最好地实现我们的各种愿望的东西与最有效地促进我们自己的长远自我利益的东西时常无法相契合。

49.自利论和道德

S和P简单地相互关联：它们两者都是有关合理性的理论。S与道德的关系更加微妙一些。一个道德理论问的不是"什么是合理的？"而是"什么是正确的？"西季威克认为，这两个问题最终是同一个问题，因为它们两者都与我们最有理由去做的是什么有关。此乃他称自我主义为"伦理学的方法"之一的原因。一个世纪之后，这两个问题看起来越来越分离了。我们已经把自我主义从伦理学中驱逐出去，而且我们现在怀疑道德地行动是"理性所要求的"。道德与自利论仍然冲突。有许多事例，其中如果某人错误地行动的话对他会更好一些。在这样的一些事例中，我们必须决定要做什么。我们必须在道德与S之间作出

选择。但是这个选择看起来是不能讨论的。每一对手的那些主张看起来与另一个对手的主张是无关的。

它们是能够被放到一起的。在那些行动理由之中，我们把道德的和利己的理由都包括进去。我们因而能够问这两种理由中哪一种更强烈一些或者具有更大的权重。正如我已经主张的，我们可能感到该问题没有什么答案。我们可能感到没有什么据以权衡这两种理由的中立的标准。但是我们并没有把该问题当作没有意义的问题而拒不予以考虑。而且不经发现一个中立的标准我们或许也能达到一个答案。我们可能发现一些能够击败自利论、表明其理由没有什么权重的论证。在第一编中，我讨论了一个这样的论证，即S是自败的这一主张。这个论证失败了。但是我将提供另一些论证，而且相信这些论证中至少有一个是成功的。

对道德理由的力度或者权重的解释将有助于这些论证。因此我们应当把对合理性的说明和行动理由包括在我们的道德理论之内。既然我们的道德理论的这一部分将关于什么是合理的，而不是什么是正确的，那么它需要比我们的理论的其他部分涉及更为广泛的范围。尤其是，它需要做可能看起来是一个错误的事情。它需要把本身并非道德理由的那些行动的理由纳入其范围内。

在这一点上那些行动者中性的理论表现得最为明显，只为所有的行动者设定共同的目标。当中立主义者们讨论道德的时候，他们可能以一种传统的方式对待自利论。他们可能把它当作一种独立的或者非道德的理论，当与道德相冲突的时候这一理论必须予以废弃，但是它具有其适当的影响领域：行动者自己的生活，限度是这并不影响到他人。但是当他们讨论合理性的时候，中立主义者们附加了自利论。通常称之为谨慎。它变得只不过是一个派生性的特殊事例。谨慎成为合理的仁爱的局部分支。这一点不但被一些功利主义者所主张，而且也为一些非功利主义者所主张，诸如内格尔。[8]

我现在能够表述另一个错误。中立主义者们附加上S可能是错误的，但是他们至少已经看到一些道德论者所忽视的一些东西。他们已经看到道德理由和利己理由可能具有共同的一些特征或根基。在这些理由并不冲突的事例中这一点最有可能。这样的一些事例因而可能是欺骗性的。在这样的事例中S可能看上去比实际上更令人信服。

最具有欺骗性的事例是这样一种事例，其中一个人的行动将只影响到他自己。我们许多人认为，在这样的一种事例中道德是缄默的。如果这个人做什么在此处不是一个道德问题，那么道德既不与S冲突也不与S契合。但在对合理性的说明中，两者在此处可能是相契合的。如果此人通过做对他来说最好的事情而遵循S，那么他将也是在做对相关的每一个人来说最好的事情。这一点是微不足道地为真的，因为他是惟一一个相关的人。但是这个真理本身并不是微不足道的。它可能引导我们得出此人是在做不偏不倚地考虑使结果最好的事情。那么S可能披上借来的外衣出现。一位自利论者可能主张，对此人而言如果以别的任何一种方式行动都会是非理性的，因为他会使结果更糟。但是S论者没有什么权利做出这一主张。根据S，对某人来说使结果更糟通常是合理的。当使结果更糟的东西也把糟糕后果转移到某个别人身上的时候就是如此。

即便我们不受这样一些事例的欺骗，也不可能有什么反对把它们束之高阁的诘难。我们可以在S与道德相冲突的那些事例中讨论S。再说一次，这一小心之举对S不可能是不公的。

50.我的第一论证

在我开始批判S之前，我将先指出基本的一点。我的一些主张可能看起来并不令人信服。即便这些主张是正确的，我们也应当有这种预期。毕竟自利论长期以来一直占据着统治地位。在长达两千多年的时间内，人们一直假定，任何一个人做他明知对他自己更糟的事都是非理性的。基督徒一直抱有这一假定，因为如果基督教为真，道德和

自利就是相契合的。如果做错事的人们知道他们将下地狱，每个人将知道，他错误地行动就是在做对他自己而言更糟的事情。基督教一直欣然诉诸自利论，因为按照他们的假设，S蕴含着作恶者是些傻瓜。类似的说法适用于伊斯兰教徒、许多佛教徒和印度教徒。既然S已经被传授了两千多年，我们必定预期在我们的直觉中会有其某种回音。S不能单凭诉诸其学说可能已经产生出的那些直觉而得到正当性的证明。

正如我在前两节中所主张的，S可能既与道德又与当前目标论相冲突。这些是能够对S加以最适当判定的事例。我的第一个论证自然而然地从这些事例的界定性特征中浮现出来。

当S与道德冲突的时候，S告诫我们每一个人给予他自己的利益以最高的权重。每个人必须由这样一个愿望所支配，那就是，他的生活对他而言要过得尽可能地好。无论给他人造成的代价是什么，每个人都必须由这一愿望所支配。我因而将称这一愿望为对自身的偏向。

我们绝大多数人有这一偏向。而且它通常比我们所有的其他愿望合到一处还要强。在这些事例中，P支持S。但我们现在假设的是这两者相冲突。我们是在考虑这样一些人，那就是，尽管知道事实而且思虑清楚，但是仍然并不想给予他们自己的自我利益以最高的权重。他们关注他们自己的利益。但这要么不是他们的最强烈的愿望，要么，如果是的话，它也被他们的其他愿望在权重上所超过。无论是哪一种情况，就这些人而言，P未能与S相契合。最好地实现他们的当前愿望的东西并不是与最好地促进他们自己的长远自我利益相同的东西。

S主张说，这些人总是应当受到对自己的偏向的支配。即便这不是他们的最强愿望，他们也应当受到这个愿望的支配。（这正是S在其并非间接自败的那些较为简单的事例中所主张的。）我们应当接受这个主张吗？

重新陈述该问题将有所帮助。批判性的当前目标论有不同的版本。一个版本与S相契合。这就是我们应当采纳的那个版本吗？在回

答这个问题的过程中，我们将更加清晰地看到接受S所要涉及的是什么。

根据CP，某些愿望可能是被合理地要求的。如果一个愿望是合理地要求的，那么我们每个人都有理由使这个愿望得到实现。即便我们并不具有这个愿望，我们也具有这一理由。如果有一个愿望被要求成为我们的终极目标，那么我们最有理由做的就是无论什么会导致该愿望得以最好实现的事情。

要使CP与S相契合，我们必须主张

> CPS：我们每个人都被合理地要求关心他自己的自我利益，而且该愿望是至高无上地合理的。对其他任何事给予同样的关心是非理性的。

按照CP的这个版本，我们每个人最有理由做的是无论什么将最好地促进他自己的自我利益的事情。

我现在可以陈述我的第一论证了。我们应当拒斥CPS。对自己的偏向并非至高无上地合理的。我们应当采纳

> （CP2）至少有一个愿望不是非理性的，而且并不比对自己的偏向更少合乎理性。当这要么是道德上可赞誉的要么是一个人的道德义务的时候，这是一个去做符合他人利益的事的愿望。

CP的这个版本与S相冲突。考虑一下

> 我的英勇牺牲。我选择我自己明知将充满痛苦的一种方式死去，但这将拯救数个人的性命。我是在知道事实且思虑清楚的情况下做我最想做而且最好地实现我的那些当前愿望的事情。（在我所举的所有例子中，这两者是相契合的。）我还知道我是在做对我而言将更糟的事情。如果我并不牺牲我的生命以拯救这些人的话，我不会遭受挥之不去的懊悔的煎熬。我的余生将活得很有价值。

按照CP的这个版本，我的行动是合理的。我之所以牺牲我的生命，是因为，尽管我关心我自己的幸存，但是我更关心他人的幸存。

根据（CP2），这个愿望不比那个对自己的偏向更少合乎理性。根据CP，鉴于该事例的其他细节，对我而言实现这个愿望是合理的。因而对我而言做明知对我更糟的事情是合理的。

51.自利论者的第一回应

在刚刚表述的那个例子中，一位自利论者必定主张说我的行动是非理性的。他因而必须拒斥（CP2）。他必须主张说我的愿望比那个对自己的偏向更少合乎理性。

S论者或许诘难说："S是一个有关合理性的理论，它并不是一个愿望理论，而是一个行动理论。我无需主张说，在你的事例中，你的愿望比那个对自己的偏向更少合乎理性。我只需主张说，既然你是在做你明知对你将更糟的事，那么你的行动就是非理性的。"

这是一个弱回应。如果S论者并不主张说我的愿望是较少合乎理性的，我们为什么应当接受他有关行动的主张？考虑一下

（CP3）如果有某个愿望（1）不是非理性的，而且（2）不比那个对自己的偏向更少合乎理性，而且（3）就某个人而言，在知道事实且思虑清楚的情况下，在考虑到所有事情的情况下，这个人最想要的就是实现他的这个愿望这一点为真，那么（4）对此人而言实现这个愿望就会是合理的。

这个主张不能够令人信服地予以否定。即便我们采纳S，我们也没有什么理由否定（CP3），因为这个主张是与S相容的。如果那个对自身的偏向是至高无上地合理的，那么S就是最好的理论。那么（除了这个偏向之外）就不会有（CP3）中所述的那类愿望。

S论者不能够令人信服地否定（CP3）。如果他对我的第一论证没有什么别的回应，他则必须提出有关不同愿望之合理性的主张。他必须主张，没有（CP3）所述的那类愿望。他必须诉诸CPS，即这样一个主张，对自身的偏向是至高无上地合理的，而且因而被合理地要求成为我们的终极目标。我将把这一对CPS的诉诸称作S论者的第一回应。

这一回应直接反驳我的第一论证。我现在将扩展这一论证。我们应当接受

（CP4）不是只有一个而是有数个愿望，它们要么并非是非理性的，要么至少不比那个对自己的偏向更少合乎理性。

考虑一下我所称的成就欲。这些是我们在工作或者更具主动性的休闲中成功地做到我们努力去做的事情的愿望。某些成就愿望可能是非理性的。比方说，就那种想要比其他任何人在山洞待更长时间的愿望而言，或者就那种想要通过行刺而达到遗臭万年的愿望而言，情况就可能如此。但是考虑一下艺术家们、作曲家们、建筑师们、作家们或者其他类型的创作者们的情形。这些人可能强烈地想要他们的作品尽可能地好。他们最强烈的愿望可能是用绘画、音乐、石料或文字创造出不朽杰作。而科学家或哲学家们可能强烈地想要作出某些基础性的发现或推进思想的进步。这些愿望并不比那个对自身的偏向更少合乎理性。我认为，这一点对其他许多愿望也为真。如果我们相信此点，那么我们的CP版本与S的冲突则更加尖锐。

值得一提的是，即使有数个并非是非理性的愿望，也可能有一个愿望是至高无上地合理的。我们不能具有可信度地就那个对自身的偏向提出这一主张。但是我们或许可以主张

CPM：我们每个人都被合理地要求去关心道德，而且这个愿望是至高无上地合乎理性的。对其他任何事情表现出同样的关心是非理性的。

按照CP的这个版本，以我们相信道德上错误的一种方式行动总归是非理性的。

CPS使CP与S相契合。类似的主张CPM却可能并不使CP与道德相契合。差别是这样的。S主张成为有关行动理由的一个完备理论，涵盖所有情形。一些道德理论提出这一相同主张。例如，此点就后果论的

理论而言就为真。但是，按照我们绝大多数人所接受的那些道德理论，道德并不提供惟一的行动理由。按照这些理论，有许多情形，其中我们可以以几种不同的方式行动，而且这些行动中没有哪一个在道德上比另一些行动更好。在这些事例中，即便我们接受CPM，我们最有理由做的也将部分取决于我们的当前愿望是什么。

52.为什么时间中立并非自利论和当前目标论之间的争议问题

就P的三个版本而言，我一直在为那个批判性的版本辩护。正如以上两节所表明的，这个版本可能与工具论的版本IP和审慎论的版本DP有很大的不同。在那许多拒斥P的学者中间，绝大多数人忽略了批判性的版本。这是不幸的。对IP和DP的许多诘难并非是对CP的诘难。在本节中我就讨论一个这样的诘难。

考虑一下那个有关我们能够合理地偏向近期的观点。我们会合理地对某个未来的痛苦关心得较少一些，不是因为它是较不确定的，而只是因为它在更远的未来。按照这个观点，合理性并不要求对一个人本身的自我利益持一种时间中立的关心。当前目标论可能看起来是这个观点的极端版本。P只是诉诸当前愿望，而且主张，要想合理地行动，我们只应当做将最好地实现这些愿望的事。P可能因而看起来是这样一种观点，那就是，要想合乎理性，我们只应当关心我们当前的利益或者我们自己的当前时刻的福利。我们称这一观点为当前自我主义（缩写为EP）。

EP的不同版本诉诸有关自我利益的不同理论。假设我们采纳的是快乐论。那么EP蕴含着

HEP：我们每人最有理由做的就是将最好地改善他当前的心态质量的任何事情。

假设我处于痛苦之中。如果我再多忍受这个痛苦一分钟，它将就此永远停止。如果我按下一个按钮，我的痛苦将即刻停止，但是几分

钟内将卷土重来，而且将再继续50年。HEP告诉我去按下这个按钮，因为通过结束我当前的痛苦，这会改善我当前的心态质量。这一改善的代价会是50年的痛苦这一点并不相干。HEP假定，不是最关心一个人当前的心态这一点是非理性的。更关心接下来50年一个人的心态是非理性的。这个观点是荒谬的。

下面回顾一下当前目标论的那个工具性的版本。这个版本主张

IP：我们每人最有理由做的是最好地实现其当前愿望的任何事。

像HEP一样，IP特别诉诸当前。但是这些观点是非常不同的。根据IP，既然没有什么愿望是非理性的，而且去做一个人所相信的将最好地实现其当前愿望的事情是合理的，那么去做任何事情都会是合理的。没有哪种行动必定是非理性的。根据HEP，既然一种行动总是合理地被要求，那么其他每一种类的行动就都是非理性的。这两个观点是相对立的两个极端。

如果我们是快乐主义者，这些说法显然表明当前目标论和当前自我主义是相当不同的观点。其次假设我们不是快乐主义者。假设我们有关自我利益的理论是非限定的愿望实现论。那么EP可能与IP相契合。

我用的是"可能"这个词，因为，如果我们不是快乐主义者，当前自我主义可能就不是一个具有可能性的观点。假设我问，"现在哪个行动会最符合我的利益？"这会自然地被当作意味着"对于现在的我来说可能的那些行动中，哪一个会最符合我的利益？"但这并不是一个当前自我主义者所问的问题。他所问的是一个新奇的问题，而且这个问题充其量只具有一种非常牵强的意义。这个问题就是，"什么最符合不是我的而是我-现在（me-now）的利益？"我们可能认为这个问题没有意义，因为我-现在这个实体不能被说成具有利益。（我可能现在对某些事情感兴趣。但是这并不相干。）如果我们并不拒斥这个问题，而且我们关于自我利益的理论是那个非限定的愿望实现论，那么我们或许

主张说，最符合我-现在的利益的东西就是将最好地实现我的当前愿望的东西。那么，根据EP，这就会是我现在最有理由做的事。这也是根据当前目标论的那个工具性的版本我最有理由做的事。

如果我们接受快乐论，那么EP是荒谬的。如果我们接受非限定的愿望-实现论，那么EP可能与P的那个工具性的版本相契合。但这并非是对P的什么诘难。我们应当接受的不是那个工具性的版本而是那个批判性的版本。

CP在时间上并不是完全中立的，因为它诉诸那个行动者的当前愿望。但是我们会给CP添加上

> （CP5）我们每人合理地被要求关心他自己的自我利益。而且这个关心应当是在时间上中立的。我们每人应当同等地关心他的生活的所有部分。但是，尽管我们全都应当具有这个关心，但是这无需是我们的主导性的关心。

如果我们添加这个主张，CP则部分地与S相契合。两个理论都同意，我们应当同等地关心我们生活的所有部分。既然它们两者都要求这种时间上中立的关心，那么这个要求就不是把S和CP的这一版本区分开来。如果我们认为对我们更遥远的未来关心得少一些是非理性的，那么这并未提供接受S而不是CP的这个版本的什么理由。[9]

在本章中我提出了我的第一论证，而且描述了S论者的第一回应。这个回应主张说，那个对自己的偏向是至高无上地合理的。这个主张得到正当性证明吗？更加关心别的诸如道德、他人的利益或者各种各样的成就会是较少合理的吗？如果正确答案是否定的，那么我的第一论证取得成功。既然我不能够证明这是正确的答案，那么这个论证就不是决定性的。但是我认为我对自己的那个偏向是至高无上地合理的这一点的否定是对的。如果有单一的另一个愿望不是更少合乎理性，正如（CP2）所主张的，那么，我们应当接受一个与S相冲突的CP版本。我相信有多个这样的愿望。

即便我们接受这样一个主张，就是，在关心我们自己的自我利益方面我们应当是时间上中立的，我们也应当拒斥S。CP能够提出这一主张。S和P之间的歧见不是有关这个主张的歧见。它关心的是，如果我们是合理的，这个关心是否应当总是成为我们的一个终极目标。

在下一章中我将提出另一个针对S的论证。这将把我的第一论证置于一个更广的语境之中。第二论证以一种更加系统的方式挑战自利论。

---

[1]Hume（1），第二卷，第三章，第三节。

[2]审慎论在Brandt（2）中最充分地得到展开。

[3]我是在效法安斯库姆和诺曼。

[4]参见Foot and Williams（5）。

[5]Nagel（1），第45页。

[6]西季威克的“自明公理”，正如舒尼文德所评说的那样，全都涉及到消除行动理由的范围和力量之上的那些任意的限度（Schneewind第300页）。即便“非理性的”这一指控只有当某个关心模式划下一条任意线的时候才得到正当证明，但这会具有广泛的蕴含。

[7]也请参见Gosling，Sen（2）和（4），and Broome（2）。

[8]Nagel（1）。

[9]就一个进一步的讨论而言，参见S·Kagan，《合理性之当前-目标论》，以及我在1986年7月《伦理学》中的回应。

第七章

诉诸完全的相对性

53.自利论者的第二回应

一位自利论者或许对我的第一论证给出另一回应。假定这个S论者接受有关自我利益的愿望实现论。这将简化他的回应。我们无需假定该S论者接受非限定的愿望实现论。他可能会接受成功论，此论只是诉诸我们有关自己的生活的愿望。按照成功论，我们的其他愿望是否得以实现并不相关。我们可以把"愿望"当作意味着"相关的愿望"。

该S论者可能拒斥我所称的他的第一回应。他或许主张

（S7）自利论无需假定对自己的那个偏向是至高无上地合理的。对你的第一论证有一种与之不同的回应。任何理由的力量都是随着时间的推移而展开。你后来将有理由去努力实现你的未来的愿望。既然你将具有这些理由，那么你现在就具有这些理由。这就是你为什么应当拒斥那个告诉你去努力实现只是你当前的那些愿望的当前目标论的原因。你最有理由做的事是将最好地实现或者能够最好地使你实现你的整个生活中的所有愿望的事。

此乃这位S论者的第二回应。它将不会成功。

54.西季威克的提法

西季威克写道：

的确，从抽象的哲学的观点来看，我看不出自我主义的原则何以应当不像普遍的原则那样受到挑战。我看不出自我主义者何以可以依据一种理由拒绝接受合理仁爱的公理，而当谨慎公理与当下的倾向相冲突时，却又不基于一种相似的根据对谨慎提出挑战。如果功利主义者不得不回答"为什么我应当为另一个人的更大幸福而牺牲我自己的幸福？"他也必定可以问自我主义者"为什么我应当为未来的更大快乐而牺牲当下的快乐？为什么我对自己未来的感觉的关心应当超过对他人的未来感觉的关心？"探究一个人追求自己的整体幸福的理由对常识来说当然是自相矛盾的。但是我不明白，那些接受极端经验心理学派的观点——尽管这些观点也被公认为同自我主义的快乐主义十分相近——的人何以能把这一要求当作荒谬的东西抛在一边。就算自我只不过是个前后连贯的现象体系，就算持久的同一的"我"如休谟及其追

随者们所说的不是一个事实而只是一个虚构，我也仍然不明白：这一感觉系列中的保持着自我的那一部分何以应当更关心它的另一部分，而不是更关心另一感觉系列？[1]

被人们大量引用的这一段文字缺乏西季威克的绝大多数行文所具有的明晰性——“纯白光”。[2]在我看来，解释是这样的。西季威克的自我主义的谨慎就是有关合理性的自利论。他的这一段文字提出了两个针对这一理论的论证。该段文字不够清楚，因为西季威克在陈述这些论证中的一个论证的时候，误入歧途。

西季威克首先主张，谨慎与合理的仁爱可能受到基于“相似的”根据的挑战。这些根据是什么？最后那两句话提示一个答案。如果只是“休谟及其追随者们”不能够摈弃对谨慎的挑战，那么，休谟有关个人同一性的观点可能就是这个挑战的“根据”。据以挑战合理的仁爱的“相似的”根据可能是某个有关个人同一性的不同的观点。

这一解释符合西季威克后来的两个主张。“否认在一个人与任何一个其他人之间存在实在的、根本的区别……将是违反常识的。如果上述说法是对的，我就看不出人们何以能证明，在为个人确定合理行动的终极目的时，这种区别并不具有根本的重要性。”[3]这些主张提示一个自利论者何以能够挑战道德论者的合理的或不偏不倚的仁爱。这个挑战能够诉诸个人之间或不同生活之间的那种区别的根本本质。如果其关联之物——各个生活的统一性——是深刻而根本的，那么生活之间的区别本身则是深刻而根本的。正如我在后面将予以论证的，这正是在有关个人同一性方面常识所信奉的。按照这个观点，我们是些不同的人、每个人具有他自己要过的生活这一点具有巨大的合乎理性的和道德上的重要意义。这就支持这样的主张：每个人的至高无上的合理目标是他自己的生活过得尽可能地好。而且，自利论者能够用这个主张挑战道德论者的不偏不倚的仁爱要求。正如西季威克所提出的，这个挑战得到有关个人同一性的常识观点的支持。

这个观点被“休谟及其追随者们”所否定。正如西季威克所写到的，休谟相信“自我只不过是一个前后连贯的现象体系……持久的同一

的'我'不是一个事实而只是一个虚构"。而且西季威克提出，休谟的观点支持对自利论的一种挑战。

所提出的这两个挑战不能够都是根据充分的。在挑战不偏不倚的仁爱方面，自利论者诉诸有关个人同一性的常识观点。自利论可能转过来遭到诉诸休谟的观点的挑战。如果常识观点为真的话，那么第一个挑战是根据充分的。如果这个观点为真，那么第二个挑战诉诸的则是一个虚假的观点。既然西季威克接受了常识观点，而且认为休谟的观点是虚假的，那么他没有展开他所提出的对自利论的挑战也就没有什么令人惊讶之处了。

休谟的观点是不充足的。但是在第三编中我将捍卫一个在相关方面步休谟之后尘的观点。而且我将主张，正如西季威克所提示的，这个观点支持针对自利论的一个论证。

在第二编的所余部分，我的目标是不同的。我将继续以并不诉诸有关个人同一性本质的任何观点的那些论证去挑战自利论。其中的一个就是西季威克的那段话所提示的第二个论证。

既然所引的这一段文字使用了模棱两可的"应当"一词，它可能看起来是关于道德的。但是，正如我所写的，它是关于我们最有理由做什么的。西季威克的"合理的仁爱公理"我们可以表达为

RB：理性要求每一个人以不偏不倚地考虑后的最大可能的幸福总量为目标。

西季威克主张说，这一点可能遭到这样的发问的挑战

（Q1）"为什么我应当为另一个人的更大幸福而牺牲我自己的幸福？"

西季威克的"谨慎公理"我们可以表述为

HS：理性要求每一个人以他自己的最大幸福为目标。

这一点能够遭到这样的发问的挑战

（Q2）“为什么我应当为未来的更大快乐而牺牲当下的快乐？为什么我对自己的未来的感觉的关心应当超过对他人的未来感觉的关心？”

我说过这两个问题可能有“相似的”根据，这是因为每一个都暗自诉诸一种有关个人同一性的观点。就像他的这段文字的结尾之处和我所引述的他后来的主张所提示的那样，这可能是西季威克头脑中所想的部分内容。但是还有更为简单的解释。（Q1）拒斥那种在不同的人之间保持不偏不倚的要求。它蕴含着，一个合理的行动者可能赋予一个特定的人——他自己——以特殊的地位。（Q2）拒斥那种在不同的时间之间保持不偏不倚的要求。它蕴含着，一个合理的行动者可能赋予一个特定的时间——当前或者行动的时刻——以特殊的地位。这两个问题之所以有“相似的”根据，是因为在一个人自己和当前之间具有类比性。这一类比提供了另一个针对自利论的论证。

55.自利论如何是不完全地相对的

这个论证可以用这样一些说法来介绍。西季威克的道德理论要求他所称的合理的仁爱。按照这个理论，一个行动者可能并不给予他自己或当前以一种特殊的地位。在既要求人称中立又要求时间中立方面，这个理论是纯粹的。另一个纯粹的理论是当前目标论，这个理论既拒斥人称中立要求又拒斥时间中立要求。自利论则不是纯粹的。它是一种混合的理论。S拒斥人称中立要求，但是却要求时间中立。S允许行动者专门突出他自己，但是却坚持认为他可能并不专门突出某个行动时刻。他不能给予他现在所想要的或所珍视的东西以特殊权重。他必须给予他的生活的所有部分或者他在所有时刻都想要的或珍视的东西以平等的权重。

西季威克可能已经看到，作为一个混合理论，S可能被指控为带有不一致性。如果行动者具有一个特殊的地位，那么为什么否定行动时刻的这个地位？我们可以诘难S说，它是不完全地相对的。

根据S，据以行动的理由可以是相对于行动者的。我有理由做将对我最好的任何事。这对我是一个理由，但对你却不是。你并没有理由去做将对我最好的任何事。

当前目标论者可能主张

（P1）如果理由能够是相对的，那么它们能够是完全地相对的：它们能够是相对于行动时刻的行动者。

这个主张可以诉诸一个人自己与当前之间的那个类比，或者说语词"我"和"现在"所指称的东西之间的那个类比。这个类比只有在形式的层面上是成立的。那些特定的时刻并不类似于那些特定的人。但是语词"我"指称某个特定的人的方式相同于语词"现在"指称某个特定时刻的方式。而且当我们每个人在决定去做什么的时候，他是在问"我现在应当做什么？"鉴于"我"和"现在"之间的那个类比，一个理论应该给予这两者以相同的对待。这就是为什么（P1）主张说一个理由只能对现在的我具有力量的原因。

"在此"与"我"也具有类比之处。当某个劝告者告诉我他如何逃离一个较少敌意的环境的时候，我或许抗议说，"但是在此我应当做什么？"如果一个人可以同时在几个地方的话，那么问"我现在应当做什么？"就会不够了。如果我现在可以在几个不同的地方的话，这个问题就不会是完全地相对的了。但是事实上无需这样的添加。主张一个理由只有对现在的我才具有力量就足够了。我们无需添加"在此"，因为，既然我是在此，我现在就不能够还在别处。[4]

当前目标论者或许提出两个更加大胆的主张。他或许说

（P2）据以行动的理由必须是完全相对的。我们应当拒斥那些其中蕴含理由可以是不完全地相对的主张。于是我们应当拒斥有关理由可以是相对于行动者的但却是时间上中立的主张。我们可以称这样的一些主张与完全的相对性不相容。

（P3）考虑一下任何一对以下述方式相联系的主张：第一个主张包含语汇"我"，但是并不包含语汇"现在"。除了的确包含"现在"之外，第二个主张就像第一个主张一样。称这样的一对主张是具有类比性的。如果第一个主张与第二个主张相冲突，那么它就与完全的相对性不相容。因而就应当予以拒斥。如果第一个主张并不与第二个主张相冲突，那么我们是否应当接受第一个主张的问题就是一个悬而未决的问题。但是，如果我们接受第一个主张，我们就也应当接受第二个主张。这是因为，如果我们接受第一个而拒斥第二个，我们的观点就是与完全的相对性不相容的。而且我们并不像我们应该是的那样，是在给予"我"和"现在"以相同的对待。

（P3）可能是不清楚的，而且一个当前目标论者为什么提出这个主张可能也是不清楚的。但是，当我应用（P3）的时候，这两者都可能变得清楚起来。

（P1）至（P3）这些主张所陈述的是我所称的诉诸完全的相对性。我相信，这是对自利论的一个强有力的诘难。

56.何以西季威克误入歧途

在我讨论这个诘难之前，提一提西季威克为什么未能看清它的力量将有所帮助。

在上文所引的那段文字之始，西季威克看来对这一点是明白的，就是对S的威胁来自P。他的第二句话是这样开始的："我看不出……当谨慎公理与当下的倾向相冲突时，却又不基于一种相似的根据对谨慎提出挑战。"这提示着会被一个当前目标论者所提出的问题：

（Q3）如果这并不是我在行动的时刻最想要的或者最珍视的，我为什么应当以我自己的最大幸福为目标。

这是一个很好的问题。但并不是西季威克后来所问的问题。他的那段文字提示着，他对诉诸完全的相对性有了一半的意识。但是他并

未完全陈述这个诉诸，而且他并未对此作进一步的讨论。我提示，他可能以下述一些方式误入了歧途：

（a）他可能要么忽略了当前目标论，要么假定了它并不是自利论的一个强劲对手。

（b）他是一个快乐主义者。在它的快乐论的形式中，S蕴含着

（S8）理性要求我以我自己的最大幸福为目标。

西季威克接受了这个主张。而且他可能已经认为，如果我们诉诸"我"和"现在"之间的类比，那些接受（S8）的人们也应当接受

HEP：理性要求我现在或当前时刻就以我自己的最大幸福为目标。

这是快乐论的当前自我主义的另一个表述方式。正如我已经说到的，这个观点是荒谬的。HEP的荒谬性可能已引导西季威克既拒斥"我"和"现在"之间的类比，又拒斥诉诸完全的相对性。

（c）这个诉诸告诉我们要拒斥S而采纳P。西季威克认为S是令人信服的。他可能没有看到，即便我们诉诸完全的相对性，我们也能够承认S是令人信服的。S能够以一系列主张表达出来。而且诉诸完全的相对性允许我们接受这些主张中的一些。我们能够接受S的那些与当前目标论相容的部分。西季威克认为S是令人信服的，而且他可能已经认为HEP是荒谬的。但是S中令人信服的东西是S的那些既与P的部分相容又与P的部分可类比的部分。而且荒谬的HEP与S的一个我们应当拒斥的部分具有可类比之处。既然如此，正如我将主张的，诉诸完全的相对性并不与西季威克的直觉尖锐冲突。

57.该诉诸应用于一个形式层面上

后果论者们拒斥那个对自己的偏向。例如，拉什多尔问道，为什么"一个不偏不倚的或者非人称的理由……应当更加重视一个人的快乐而不是另一个人的快乐？"他承认"从一个人自己的观点来看应该合理地欲求是一件事，而当他接受一个更大的整体的观点的时候则是另一件事……是可理解的。[5]但是这两种观点都是同等合理的吗？一旦那人知道整体的存在……秉持部分的观点如何能够是合理的？"[6]

一个自利论者可以回应道："我们不是在问理由去做什么是合理的。我们是在问我去做什么是合理的。而且我可能合理地拒绝接受'整体的观点'。我不是那个整体。我的观点为什么不可能恰好就是我的观点？"

自利论可能反过来遭到挑战。一个当前目标论者可以说："我们不是在仅仅问我做什么是合理的。我们是在问我现在做什么是合理的。我们必须考虑的不仅仅是我的观点，而且是我当前的观点。"正如威廉斯所写到的，"对一个人的生活的正确视角来自现在。"[7]

这一点能够以更为正式的说法加以表达。步内格尔之后尘，我区分了两种行动理由。如果一个理由并不受任何一种观点的束缚，内格尔则称之为客观的。假设我们主张有理由解除某个人的苦难。如果这个理由是一个对于每一个人来说的理由——对任何一个会解除这个人的苦难来说的理由，那么这个理由是客观的。我称这样的理由为对行动者中立的。内格尔的主观的理由只是对该行动者来说的理由。我称这些是相对于行动者的。[8]

我应当进一步解释一下在什么意义上理由能够是相对的。在一种意义上，所有的理由都能够是相对于一个行动者、一个时间和一个地点的。即便你我都是在努力达到某个共同的目标，但我们仍可能处于不同的因果境遇之中。我可能有一个以促进我们的共同目标的方式行动的理由，但是你可能没有这样的理由，因为你可能不能以这一方式行动。既然甚至对行动者中立的理由在这个意义上也能够是相对于行动者的，那么这个意义对这个讨论来说就是不相干的。

当我称某个理由是对行动者中立的时候，我并不是主张说，这个理由对其他行动者来说不能够是一个理由。我所主张的无非是它可能不是。按照当前目标论，如果其他那些行动者与我具有相同的目标，我的行动理由对他们来说也是一个理由。同样，当P主张说某个理由是相对于时间的时候，该主张并不是说，随着时间的流逝，这个理由注定会失去。该主张只是说它可能失去。如果该行动者的目标出现变化的时候，它就将失去。

如果所有行动理由都是对行动者中立的，这对自利论会是致命的一击。考虑一下每一个人促进他自己利益的理由。如果这对每一个人都是一个理由的话，每一个人都会具有促进每一个人的利益的同等理由。自利论就会被不偏不倚的仁爱所吞并。因而自利论者必须主张说，行动理由能够是相对于行动者的。它们能够是行动者的理由，而不是其他任何人的理由。

一位当前目标论者会予以赞同。但是他能够添加（P1）：即一个理由能够是相对于行动时刻的行动者的这样一个主张。在别的那些时刻它对该行动者不是一个理由的情况下，它能够在这个时刻对该行动者来说是一个理由。

这个主张对S论者的第二回应构成挑战，后者假定任何一个理由的力量都是历时地展开的。根据S，正如内格尔所写到的，“有理由促进……将有理由的东西。”[9]利己的理由在这个意义上是无时间的或时间上中立的。

尽管是无时间的，但是它们并非是非人称的。正如我所主张的，S是一个混合理论。根据中立论者的道德理论，那些行动理由既是无时间的又是非人称的。根据当前目标论，那些理由既是相对于时间的又是相对于行动者的：它们是该行动者在行动时刻的理由。根据自利论，那些理由是相对于行动者的，但不是相对于时间的。尽管S拒斥非人称性这一要求，但是它却要求时间上的中立。

作为一个混合理论，S能够受到两个方面的夹击。而且S针对一个对手所持的主张，另一个对手可以拿来针对它。在拒斥中立论方面，一个自利论者必须主张一个理由只有对该行动者而言才可能具有力量。但是据以提出这一主张的那些根据支持一个进一步的主张。如果一个理由只有对该行动者而言才能够具有力量，那么它只在行动的时刻才对行动者具有力量。该自利论者必须拒斥这个主张。他必须对相对于时间的理由这样一个概念发起攻击。但是，用以表明理由必须是对时间中立的这一点并从而驳倒当前目标论的那些论证，可能也表明理由在不同的人之间必须是中立的并从而驳倒自利论。

内格尔一度提出过属于这第二类的一个论证。如果他的论证成功，中立论则获胜。我现在所讨论的不是内格尔的论证，而是在讨论诉诸完全的相对性。像内格尔的论证一样，这个诉诸挑战自利论。这个诉诸的一个部分是（P1），即这样一个主张：如果理由能够是相对于行动者的，那么它们能够是完全相对的——相对于行动时刻的行动者。

理由要么能够是相对的，要么不能够。如果它们不能够，正如内格尔所论证的，那么中立论获胜。我们不但既要拒斥自利论，又要拒斥当前目标论，还要拒斥绝大部分常识道德。

下面假设，正如内格尔现在所相信的，理由能够是相对的。（P1）正确地主张，如果理由能够是相对的，那么它们能够是相对于行动时的行动者的。正如我将在下面一章中论证的，在我现在具有这个理由这一点不为真的情况下，我一度具有一个促进某个目标的理由这一点会是真的。而且在我现在具有这个理由这一点不为真的情况下，我将具有一个促进某个目标的理由这一点会是真的。既然这两者都会是真的，那么就不会主张说，任何理由的力量都是历时地展开的。这从根本上削弱了S论者的第二回应。

除了诉诸（P1）之外，一个当前目标论者能够诉诸更大胆的（P2）和（P3）。我现在将表明，如果这个诉诸得到正当的证明，那

么就存在据以拒斥S的一些进一步的根据。

58.该诉诸应用于其他一些主张

首先假设某个人既接受S又接受有关自我利益的愿望实现理论。当S应用于他自己的时候，这样的一个人会用这样的主张来陈述S

（S9）我最有理由做的是将最好地实现我的整个生活中的所有愿望的事。

与之具有类比性的主张是

（P4）我现在最有理由做的是将最好地实现我现在所具有的所有愿望的事。

既然这些主张是相互冲突的，那么诉诸完全的相对性告诉我们要拒斥（S9），后者与完全的相对性是不相容的。该诉诸告诫我们拒斥这个主张，那就是告诫我们拒斥S的一个版本。而且它允许我们接受（P4），它陈述的是P的那个工具论的版本。

其次考虑一下

（S10）我能够合理地忽略不是我的愿望的那些愿望。

而且

（P5）我现在能够合理地忽略现在不是我的愿望的那些愿望。

这些主张并不相互冲突。根据诉诸完全的相对性，我们因而应当要么接受两者要么不接受其中的任何一个。既然一个自利论者必须接受（S10），那么它必须接受（P5）。但是（P5）既是对S的一个否定，又是对P的一个部分陈述。如前所述，该诉诸算来不利于S而有利于P。

尽管一个当前目标论者会拒斥自利论，但是他会接受S所做出的某些主张。于是他会接受对RB的拒斥，后者是这样一个主张，那就是理性要求不偏不倚的仁爱。而且，尽管他会拒斥

HS：理性要求每个人以他自己的最大幸福为目标。

但是他会代之以接受

（P6）更加关心一个人自己的幸福并非是非理性的。

如果他不是一个快乐主义者，他应当添加

（P7）更加关心发生在自己身上的事或者更加关心一个人自己的自我利益并非是非理性的。

这个主张为我所称的对自身的偏向作了辩护。但是与S不同的是，它并不坚持认为，一个合理的行动者必须既具有这个特定的偏向又受制于这个特定的偏向。它只是主张这个偏向并不是非理性的。

假设我们接受（P7）。与之具有类比性的主张是

（P8）更加关心当前时刻发生在一个人自己身上的事并不是非理性的。

这个主张为我将要称作的当前偏向作了辩护。这一偏向甚至比那个对自身的偏向更加常见。这两个偏向可能以这样一些思想表达出来：“我知道某人不得不去做这个糟透了的工作，但是我希望那不是我”；“我知道我不得不迟早去做这个工作，但是我希望我不是现在做这个工作”。（或者：“我知道我的牙齿得要钻洞修补，但是我希望不是就在这一个时刻钻洞修补”。）

根据诉诸完全的相对性，如果我们接受（P7）我们也应当接受（P8）。这看来是正确的。这两个偏向可能根据相似的根据予以辩

护。我的那些关心发生在我身上的事情的理由不同于我的那些关心发生在他人身上的事情的理由。我与我的感受之间的关系比我与他人的感受之间的关系要密切得多。这个事实使（P7）令人信服。同样，我的那些关心在我身上现在发生的事情的理由根本不同于我的那些关心在我身上已经发生或者将会发生的事情的理由。现在的我与我现在的感受的关系比现在的我与我在其他那些时候的感受之间的关系要密切得多。这个事实使（P8）令人信服。

既然（P7）既可能是自利论者的主张，又可能是当前目标论者的主张，那么后者能够说："自利论部分地是正确的。其错误在于从（P7）转移到一个更加大胆的主张。根据（P7），更加关心一个人自己的自我利益并非是非理性的。这是令人信服的。但是主张说，如果我们是合理的，对我们自己的自我利益的关心必须总是我们的一个终极目标这一点则并不令人信服。"

（P7）和（P8）对当前目标论而言并不是主要的。如果某人既不对他自己也不对他自己的当前感受更加关心的话，他不会被当前目标论者判定为非理性的。这两个主张并不为那个与之相当的不同的主张所蕴含，该主张对于P的批判论的版本而言是主要的。（该主张是说，如果我们知道那些事实而且思虑清楚，那么我们最有理由做的是任何会最好地实现我当前的愿望中的并非非理性的那些愿望的事。）

尽管一个当前目标论者会接受（P7），但是他会把它嵌入一个更大的主张之中。这或许是

（P9）一个关心模式并非仅因为它并不给予获得不偏不倚地考虑下来的最好的结果以至高无上的权重而就是非理性的。更加关心发生在我身上的事、更加关心我所爱的人们、更加关心我所努力达到的成功或者我所致力的事业，对我来说并不更少合乎理性。

这个主张并未给予那个对自身的偏向以什么特殊的地位。我援引这个偏向只是作为一个例子，用以说明一个关心尽管不是不偏不倚的或者不是对行动者中立的，但是并不更少合乎理性。这是最简单的而

且是偏向最明显的或相对于行动者的关心。但是还有无数的其他关心，（P9）就提到其中的一些，并且主张它们并不更少合乎理性。

如果我们接受（P9），那么诉诸完全的相对性告诉我们要接受

> （P10）一个关心模式并不仅因为它并不给予一个人自己的自我利益以至高无上的权重而就是非理性的。现在更加关心我现在所爱的那些人们、更加关心我现在所努力达到的成功或者我现在所致力的事业，对我来说并不更少合乎理性。而且，我对我自己的自我利益的关心涉及到相对于当前的一个时间上的偏向可能并不更少合乎理性。例如，就我来说现在更加关心在我身上现在所发生的事这一点可能并不更少合乎理性。

（P10）是当前目标论的主要主张之一。它并不给予当前偏向以什么特殊地位。这只是作为一个例子引用，用以说明一个关心尽管是相对于行动时刻的行动者的，但是可能并不更少合乎理性。

我使用了"可能"一词，是因为CP有不同的一些版本。其中一个版本主张，在我对我自己的自我利益的关心方面，我应当是时间上中立的。这与（P10）相一致。按照CP的这个版本，对于那些我所爱的人的利益我也应当具有时间上中立的一种关心。在努力做对于这些人而言最好的事情方面，我应当给予他们的生活的所有部分以同等的权重。这符合就我而言现在更加关心我现在所爱的那些人的利益并不更少合乎理性这一主张。而且，即便我对我自己的自我利益的关心应当是时间上中立的，就我而言更加关心我现在所努力达到的成功或者我现在所致力的事业这一点也并不更少合乎理性。我对这一成功或者这些事业的关心并不是对我自己的自我利益的关心。

如果我们接受（P10），我们将拒斥自利论的那个主要主张：对任何一个人而言做他明知将对他自己更糟的事情是非理性的。根据S，一个至高无上地合乎理性的关心模式是一种时间上中立的对自身的偏向。根据（P10），其他许多关心模式并不更少合乎理性。如果是这样，我的第一论证取得成功。如果我们具有这些其他关心模式之一，那么，按照它采取行动并不会更少合乎理性。甚至当如此行动时

我们会是在做我们明知有悖于我们自己的长远自我利益的事，这也不会更少合乎理性。

我们应当接受（P10）吗？更确切些说，如果我们已经接受（P7）——对自身的偏向并不是非理性的这一主张的话，我们能够拒斥（P10）吗？这个偏向是独一无二地或者至高无上地合乎理性的吗？我已经提到过别的一些我主张的并不更少合乎理性的愿望和关心。如果那一个主张证明具有正当性，我们应当接受（P10）的一个被限定的版本。

我们之所以应当限定（P10）是因为具有某些想要取得成就的愿望或致力于某些事业可能是非理性的。于是，想比其他任何人在山洞中待更长的时间可能是非理性的。但是，正如我已经主张的，有许多想要取得成就的愿望并不比那个对自身的偏向更少合乎理性。想要创造出某些种类的美或者取得某些种类的知识并不更少合乎理性。而且还有许多别的例子。类似的主张适用于我们所致力的那些事业。在19世纪致力于采用世界语为全世界的交际语或曰国际性语言并不是非理性的。鉴于英语和世界语所处的相对地位，现在再致力于这种努力可能就是非理性的。但是还有许多别的事业，致力于这些事业并不比那个对自身的偏向更少合乎理性。既然我们应当接受（P10）的一个被限定的版本，我们就应当接受CP的一个与S时常冲突的版本。

在本章中我论证了，如果一个理由只能对一个人具有力量，那么一个理由对一个人只能在某个时刻具有力量。我们应当拒斥有关任何理由的力量都是随着时间的推移而展开的主张。我们因而应当拒斥那个诉诸该主张的、回应我的第一论证的S论者的第二回应。如果S论者没有什么别的回应，他可能不得不回归他的第一回应。他可能不得不主张，那个对自身的偏向是至高无上地合乎理性的。我们应当拒斥这个主张。如果该S论者没有什么别的回应，我们就应当拒斥S。

我还论证说，拒斥S还有别的一些根据。这些是由诉诸完全的相对性所提供的。根据这一诉诸，站得住脚的理论惟有道德论和当前目

标论，因为只有这些理论给予“我”和“现在”相同的对待。（那些对行动者中立的道德理论显然给予“我”和“现在”以相同的对待。尽管较不那么明显，那些相对于行动者的理论也是如此。这些要求我给予某些人的利益以特殊的权重。于是我应当给予我的子女的利益以特殊的权重。这一主张可能看起来给予“我”一种待遇，却剥夺了“现在”享受这一待遇。但是情况并非如此。在某个时候我与我的子女的关系不能成立，在某另一个时刻则未能成立。我的子女本来不可能既存在又不是我的子女。这就是为什么，在那个关乎我对我子女的义务的主张中，我们无需包括“现在”一词的原因。还有一些关系在某个时刻能够成立而在某些别的时刻无法成立。一个相对于行动者的道德理论在其有关这些关系的那些主张中，的确包括了“现在”一词。例如，如果我是一个医生，我对那些现在是我病人的人们具有特殊的义务。我对那些一度是我的病人而现在是某个别的医生的病人的人们则不具有这样的义务。）

在下一章中我提供了一些据以拒斥S的进一步的根据。但是我的主要目标是要讨论一些令人迷惑不解的问题。

---

[1]Sidgwick（1），第418—419页。

[2]Blanshard。

[3]Sidgwick（1），第498页。

[4]在第87节中所讨论的那个想象出的、我的头脑一分为二的事例中，我的理由会需要是三重相对的。我会需要问：“在我的这一半头脑之中我现在应当做什么？”

[5]G·E·摩尔例外。“西季威克教授用‘对他自己而言的那个终极合理目的’和‘对他首要的’这些短语意指什么？他并未尝试对它们加以界定；而且在很大程度上正是这样一些未界定的短语的使用造成哲学中所犯的那些荒谬。”（Moore第99页）摩尔对这些短语有意义这一点的否定构成他对西季威克的“驳斥”。他补充说：“不会再有对任何理论进行完全而彻底驳斥的愿望了。”摩尔本人刚刚主张“善”不能够定义。正如麦基所发出的公正惊呼（Mackie（1），第323页），“多么厚颜无耻！”

[6]Rashdall，第45页。

[7]Williams（3），第209页。

[8]Nagal（1）和（3）。

[9]Nagal（1），第45页（斜体是我所加）。

第八章

对时间的不同态度

自利论主张，在我们对自己的自我利益的关心方面，我们应当是时间上中立的。正如我已经说到的，一个当前目标论者也能够持有这一个主张。我现在将要问的是，这个主张是否得到了正当性的证明。如果答案是肯定的，这就不是什么对P的诘难。但是，如果答案是否定的，这就是对S的另一个诘难。

59.不给一个人过去的愿望以权重是不合理的吗？

首先考虑一下那些接受有关自我利益的愿望实现论的自利论者。按照这个理论的所有版本，对某人而言最好的东西就是将最好地实现其整个生活中的所有愿望的东西。而且某人愿望的实现对他而言是好的，而不能实现其愿望对他而言则是糟的，即便这个人永远不知道这些愿望是否已经得到实现。

在决定什么会最好地实现我的愿望方面，我们必须力求预测，如果我的生活是以那些可能的不同的方式进行的话我们会有什么样的愿望。如果一个愿望更强一些，那么该愿望的实现的重要性要更大一些。如果我具有这个愿望的时间更长一些的话，它也应当算作更重要一些吗？就那些强烈愿望的事例而言，肯定的回答看来是令人信服的；但是就那些弱愿望而言答案就不那么明显了。

按照非限定愿望实现论，如果我的愿望中的任何愿望得到实现对我而言则都是好的，如果任何愿望未能实现对我而言则是糟的。另一个版本是成功论。这一理论只是赋予关于我自己的生活的那些愿望以权重。这些愿望是哪一些并不总是清楚的。但是这决不是对成功论的什么诘难。对我而言什么会是最好的这一点为什么总是应当清楚明白？

其次我们可以回顾一下成功论如何不同于快乐论的那个最为广泛的版本。两个理论都诉诸一个人的有关他自己的生活的愿望。但是快乐主义者只是诉诸与我们生活中的那些内省地可觉察到的特征相关的愿望。假设我的最强愿望是解决某个科学难题。快乐主义者主张，如果我在我的余生认为我是在解决这个难题，那么对我而言将会更好。按照他们的观点，如果我的信念是虚假的，这一点并不重要，因为这对所经验到的我的生活质量并不造成什么差异。明白与误信并非不同的体验。按照成功论，如果我的信念是虚假的，那么对我而言将会更糟。我所想要的是解决这个难题。如果这个愿望得不到实现，即便我相信这个愿望是虚假的，对我而言也将会更糟。

最后谨记快乐论和成功论两者都有两个版本。一个版本所诉诸的是一个人的那些局部的或特定的愿望的实现总量。另一个版本所诉诸的只是一个人的那些全局性的愿望：那些要么有关他的生活的诸部分，要么有关他的整个生活的优先选择。我或许全局性地宁取两个可能的生活中的一个，即便这个生活包含较小一些局部性的愿望实现总量。那一对在他们迷狂出神的巅峰跳下悬崖的夫妇就是一个这样的全局性优先选择的例子。

我们本来能够区分出愿望实现论的一些别的版本。但是在此并不必要。我将挑战那些假定愿望实现论的某个版本的自利论者。我将再次使用"愿望"来指"相关的愿望"，因为这个理论的不同版本诉诸不同的愿望。无论假设的是哪一个版本，我的绝大多数说法都会适用。

按照愿望实现论，西季威克的合理的仁爱公理变成

（RB1）每一个人最有理由做的事就是将最好地实现每人的愿望的事。

而且自利论变成

（S11）每一个人最有理由做的事就是将最好地实现或能够最好地使他实现他自己的所有愿望的事。

一个自利论者必须拒斥（RB1）。正如我所说的，他可能主张

（S10）我能够合理地忽视不是我的愿望的那些愿望。

一个当前目标论者可能会增加考虑

（P5）我现在能够忽视现在不是我的愿望的那些愿望。

[我使用“可能”，是因为按照P的那个批判论的版本，（P5）或许遭到拒斥。] 根据诉诸完全的相对性，如果我们接受（S10），那么我们也应当接受（P5）。既然一个S论者必须接受（S10），但是不能够接受（P5），那么他必须拒斥诉诸完全的相对性。该诉诸主张，理由能够是相对的，不仅相对于特定的人们，而且相对于特定的时间。S论者或许回应说，尽管何人具有某个愿望这一问题具有巨大的合理的意义，但是何时该愿望被具有这一问题则不具有这样的意义。

情况是这样的吗？我应当力求实现我过去的愿望吗？有关愿望实现论也能够问相似的问题。我过去的愿望的实现对我而言是好的，而这些愿望无法实现对我而言是糟的吗？除了死者的遗嘱之外，过去的愿望很少被愿望实现论讨论到。这可能是因为那个具有类比性的问题按照更老的快乐论的某些版本不能够出现。我现在不能够改进我的过去经验的质量。但是，即便当我不再想这么做的时候，我也或许能够实现我过去的愿望。我有这么做的理由吗？

一些愿望在它们自己的持续性方面是暗含地有条件的。如果我现在想要到月亮升起以后的时候去游泳，那么只有当月亮升起时我仍然想要去游泳的时候，我才可能想要这么做。如果一个愿望在其持续性方面是有条件的，一旦它成为过去，那么它显然是能够被忽略的。

有一类愿望，其中许多就是以这种方式暗含地有条件的。这些是我们相信其实现会带给我们满足或其未能实现会带给我们苦恼的一些愿望。在我们不再具有这些愿望之后，两种情况中没有哪一个会发

生。就许多这些愿望而言，它们自己的持续性是有条件的，因而这一点是自然而然的事。

就别的愿望的情况而言，并没有这样的一般性理由使之如此。假设我在火车上遇到某个陌生人。她对她一生的抱负以及她对自己成功的机会所抱有的希望和担心表白了一番。在我们旅途的终点，我的同情心油然而生，强烈希望这个陌生人能够成功。尽管我知道我们永远不会再相遇，但我仍然具有这一强烈的愿望。这个愿望在其持续性方面并不是暗含地有条件的。就许多别的愿望而言同样为真，就许多种类的愿望而言同样为真。

最为清楚的一些例子是某些人所具有的有关死后将发生什么的愿望。假设我并不相信我将有来生。既然我相信我的死亡将是我的终结，我的那些有关在这之后会发生什么的愿望就它们持续到实现的那一刻的持续性而言不可能是有条件的。我相信这个条件不会得到实现，但是我仍然具有这些愿望。而且当我死亡的时候，这些愿望将成为过去的愿望，它们在其持续性方面并不是有条件的。这些——死者的无条件愿望——并不使一些愿望实现论者感到尴尬。他们接受这样的主张，即，在导致这样的愿望未能实现方面，我们的行动有悖于死者的利益。

如果他们假定成功论，他们也不会对有关所有这些愿望都提出了这一主张。我的最强烈的愿望之一是威尼斯城永不被毁坏。假设当我死了之后，潮水摧毁了威尼斯城。按照成功论，这不会有悖于我的利益或蕴含着我具有一个更糟的生活。但是成功论把某些事件看作对于某个死去的人来说是件糟糕的事。假设我五十年如一日地致力于确保威尼斯得到保全。按照成功论，如果当我死了之后，威尼斯遭到毁灭，那么对我而言将会更糟。这将使这样的一点为真，就是我毕生的工作白费了。这将使我的一生在我最为在乎的一个方面失败了。当威尼斯遭到毁灭的时候，我们应当主张，我的生活不如我们原先认为的那样好。

我们应当接受最后这个主张吗？它看起来是可以得到辩护的，但是对它的否定也是如此。如果我们否定这个主张，我们看起来是在诉诸快乐论。我们看起来是在假定，如果我永远不知情的话，我毕生的工作白费这一点对我而言并不能够是糟糕的。看清我们为什么应当相信这一点是困难的，除非我们假定，如果一个事件对我所经历到的我的生活质量没有关系，那么它对我而言就不能是糟糕的。

尽管死者的愿望在道德上是有趣的，但是它们与本讨论并不相关。我是在考虑那些假定愿望实现论的某个版本的那些快乐主义者。根据这些人，我们每个人最有理由做的事乃是将最好地实现或能使他最好地实现他的整个生活中的所有愿望的事。我不能够问，如果一个人死了他是否应当力求实现他过去的愿望。我必须问我的一个有关一个活着的人的过去愿望的问题。这些必定是这样一些愿望，尽管它们没有持续存在，但在它们自己的持续性方面并不是有条件的。

我们可以把同一个例子改变一下。假设，五十年来我不但致力于保全威尼斯城，而且定期向威尼斯保护基金会捐钱。在这整个五十年间，我最强烈的两大愿望就是威尼斯得以保全和我成为保全它的人之一。这些愿望在其持续性方面并不是有条件的。我想要威尼斯得以保全而且我自己成为其保全者中的一员，即便我后来不再具有这些愿望。

其次假设我的确不再具有这些愿望。因为我在建筑学方面的趣味发生变化，我不再关心该城的命运。我仍然有理由向威尼斯基金会捐款吗？按照S的那个诉诸愿望实现的版本而言我具有这样的一个理由吗？我有理由不再捐款，因为用所省下来的那些钱我本来能够实现我的一些当前的或者未来的愿望。我具有一个与之相反的理由而去继续捐款吗？如果我进一步捐款，这可能有助于实现我的两个过去的愿望。尽管我不再具有这些愿望，但是五十年来它们一直是我的最强烈的愿望。

那个更广泛的问题是这样一个问题。按照愿望实现论，我应当给予我的所有愿望——不论它们是过去的、现在的还是将来的——以同等权重吗？当决定什么对我而言是最好的时候我应当给予所有这些愿望以同等权重吗？（我是指“其他事情同等的情况下，同等的权重”。如果两个愿望中一个更弱一些，或者我后悔有这样的一个愿望，抑或某些别的主张为真，那么我应当赋予这个愿望以较小一些权重。但是，因为我的一些愿望不是我当前的愿望我就应当给予它们较小一些权重吗？）

假设一个S论者给出否定回答。而且假设他还主张，至少就非常强烈的那些愿望的情况而言，如果它们延续更长久的话，它们的实现就更加重要。如果我不再捐款的话，这会使我能够实现我当前和未来的一些愿望。但是我只会再活一些年而已。于是这样一点可能为真，那就是，如果我们既计算愿望的强度又计算愿望的持续时间，那么我现在的和未来的那些愿望合起来，其重要性也会比五十年来我两个最为强烈的愿望要小。如果我的整个生活中的我的所有愿望都应当得到同等权重，那么S论者可能不得不得出这样的结论，那就是，如果我现在不再捐款，对我而言会更糟。那么他必定主张，对我而言这么做会是非理性的。即便我现在并不具有、将来永远不会具有任何捐献的愿望，不再捐款也会是非理性的。

这个结论可能令自利论者尴尬。他可能受到诱惑，得出一个合理的行动者能够忽略他过去的愿望的结论。但是，如果他据以得出这个结论的根据是，这些愿望是过去的，这可能是一个损害性的让步。那么S论者必须放弃这样一个主张，那就是，当某个愿望是糟糕的时候，它不能够是合理地具有重要意义的。但是他仍然必须主张，一个合理的行动者应当给予他当前的和未来的那些愿望以同等权重。而且，如果这个主张不能够得到一种诉诸时间中性的支持，它可能更加难以辩护。

60.取决于价值判断或者理想的那些愿望

如果S论者想要诉诸时间中性，他就必须提供别的理由，以说明为什么我们能够忽略我们过去的愿望。他或许诉诸我们能够失去某个愿望的那些方式之一。我们能够改变我们的主意。在某种意义上，我们的愿望中的任何变化都涉及主意的改变。这里所说的是某个价值判断或者理想方面的变化：有关值得欲求的是什么的观点的变化。

我对什么是我们最有理由去做的与鉴于我的信念什么由我来做会是合理的作了区分。如果我的酒中被人投了毒，喝下这样的毒酒不是我最有理由去做的。但是，如果我没有什么理由相信我的酒被人投了毒，那么我倘若喝下这杯酒就不会是非理性的行动。我的主要问题是关于我们最有理由去做的是什么。但下文则是关于在某人的信念已定的情况下对他来说做什么是合理的。一些人称之为主观上合理的是什么的问题。

S论者可能说："任何一个人都能够合理地忽略他所丧失掉的愿望，因为他改变了主意。而且，当你不再想要威尼斯得到保全的时候，你必定已经改变了你的主意。与之形成对比的是，当某人死亡的时候，他的愿望成为过去的愿望，但是并没有改变主意。这就是为什么当我们考虑某个人的利益的时候不应当忽略他弥留之际所具有的愿望的原因。"

这不可能是一个完整的回答，因为它只涵盖依赖于价值判断或者理想的那些愿望。我们具有许多更为简单的愿望，而它们的丧失并不涉及主意的改变。

当这个回答应用于其中我们并不改变我们的主意的那些事例的时候，这是一个好的回答吗？在后面讨论到的少数事例中，我们把我们的主意的改变当作一种讹误。在所有其他事例中，持有下述这样的主张是令人信服的，那就是，当某人因为改变了主意而丧失一个愿望的时候，他能够合理地忽略这个愿望。这是因为该愿望依赖于他现在所拒斥的一个价值判断或者理想。

尽管这一主张令人信服，但是在为S辩护方面却不是一个好的回答。相似的主张适用于那些将被以后某个主意的改变所产生出的愿望。假设我预计以后将具有一些将依赖于我现在所拒斥的一些价值判断或者理想的愿望。我应当给予这些未来的愿望与我当前的那些愿望以相同的权重吗？S论者必定给出肯定的回答。

内格尔举出一个相关的例子：

> 假设（某人）现在相信20年后他将珍视安全、地位、财富和安宁，而他现在所珍视的是性爱、自主、经常冒险和强烈感受。对这种境况的一个决定性的反应会呈现两种形式之一。该个体可能足够坚定地确信他的这些终将出现的未来价值其实毫无价值，以至于现在干脆拒不予以关心……另一方面，他可能把他目前的价值和他的这些未来的价值都当作优先选择来对待，把每一个都当作一个更高的原则——“按照自己选择的方式生活”——之下的一个理由源泉。那会要求他具有一种谨慎，保持通往最后的受尊重之路的畅通。[1]

正如内格尔所写的，S要求这个年轻人把他的价值和理想当作它们宛如纯粹的优先选择来对待。只有如此，对他而言给予他的那些预见到的未来价值以同等权重才会是合理的。S主张，他千万不能以那些他将预见到会后悔的方式行动。他千万不能支持政治运动或者签名请愿，倘若这有可能严重地妨碍或者限制他比较保守的中年自我的机会的话。更确切地说，如果不偏不倚地算计出来的结果不错的话，他可以以这些方式行动。在这种算计中，他可能会因为他以后将有的各种各样的不同价值或理想的可能性较小而打折扣，但是却千万不能仅仅因为他现在认为它们毫无价值或者格格不入而对这样的预见到的价值或者理想大打折扣。

如果我们认为我们的价值或者理想不是纯粹的优先选择，而且能够或多或少地得到正当性的证明，我们就不能够接受最后这个主张。我们不能够完全接受S。我们必须像P所主张我们所应当做的那样对待我们的价值或者理想。我们必须给予我们现在认为已得到了较好证明的东西以特殊地位。

如果我们持有基于道德信念的那些愿望，而不是像内格尔的例子中的那些愿望，那么上述这一点甚至更加明显。即使我们假定这些信念不可能或多或少地得到正当性的证明，而是直接为真，它仍然是更加明显的。那么这一点就所有信念而言是常见的。我们不能够正正当当地说：“Q是真的，但是我现在并不相信Q。”我们现在必须认为我们现在所认为为真的东西为真。但是这个有关我们当前信念的主张并不涵盖我们过去的或者未来的信念。我们能够正正当当地说，“Q是真的，但是我一度并不相信Q，而且我将来可能不再相信Q。”

有关可估价值的那些愿望的相应的一点则需要小心陈述。内格尔写道：“该个体可能足够坚定地确信他的这些终将出现的未来价值其实毫无价值，以至于现在干脆拒不予以关心。那么他会把他当前的价值当作对未来也是有效的来看待，而且从他的那些预期到的未来观点中不会衍生出什么谨慎的理由。”内格尔还补充说，如果这就是该个体的反应，那么“他的地位按照那些无时间性的理由衡量会是可公式化的。”他的那些理由会以一种无时间性的方式得到陈述，因为它们会诉诸他所相信的其有效性无时间性的那些价值。虽然如此，在另一个意义上，他的那些理由既会是相对于时间的又会是相对于行动者的。我们议题中的这个人是在努力按照这样一些价值在行动，它们的有效性是无时间性的。但他充其量所能做的就是他按照他现在所认为有效的那些价值去行动。这正是当前目标论所要求他做的。

按照自利论，这个年轻人必须给予他当前的与他的预见到的未来价值和理想以相同的权重。这会是在给予他现在所认为得到正当性的证明的东西与他现在认为毫无价值或者不值一提的东西以相同的权重。这显然是非理性的。它甚至可能在逻辑上是不可能的。

我们已经得出一个一般性的结论。根据S，我应当给予我所有的当前愿望与未来愿望以同等权重。这个主张甚至适用于将依赖于我们的价值判断或者理想方面的变化的那些未来愿望。当被应用于这些愿望的时候，这个主张是无法辩护的。就那些基于价值判断或者理想的

行动理由而言，一个合理的行动者必须给予他现在所接受的价值或者理想以优先权。就这些理由而言，正确的理论不是S而是P。

这个结论还有一些进一步的根据。假设我认为，随着知识和经验的增长，我将变得更加具有智慧。按照这个假设，我应当给予我的未来的可估的愿望比我当前具有的可估的愿望以更大的权重，因为我未来的愿望将更好地得到正当性的证明。该主张可能看起来与P相冲突。但是情况并非如此。如果我既假定我总是会越来越具有智慧，又假定现在能够预见在主意方面未来某种特定的变化，那么我就实际上已经改变了我的主意。如果我现在相信某个后来的信念将更好地得到正当性的证明，那么我现在就应当具有这个信念。这样，我变得越来越具有智慧这一假定并不提供对P的什么诘难。甚至按照这个假定，我仍然能够给予我现在认为得到正当性证明的东西以特殊地位。

相反的假定是，随着时间的推移，我将变得更少具有智慧，而且我的理想方面出现的变化将是一个讹误。理想的丧失是稀松平常的事；而且判断力也时常出错。（在许多诗人那些相继的《诗选》中，所选入的诗作越来越差。）按照这个假定，我应当给予我一度所珍视或所相信的东西以更大的权重。但是同样，如果接受这个假定，我仍然会相信我一度所相信的，即便我不再那么在乎。

就这些相反的假定而言，没有哪一个看起来在一般的意义上更有可能得到正当性的证明。既然这些假定相冲突，作为一种妥协，我们可以提出时间中性是自利论所要求的。而且有一个怀疑论的论证可能看起来有利于这样的中立。我可能受有关我当前确定性的专横推断的影响。我凭什么应当假定我现在更有可能是对的？

尽管这个论证支持转移到时间中立，但是它无助于自利论者。像其他论证一样，它把我们带到他的理论之外。我凭什么应当假定我现在更有可能是正确的？怀疑论的论证既对我按照我现在所接受的价值观或者理想行动的信心形成挑战，又对我按照我所接受的价值观或者

理想行动的信心形成挑战。该论证支持给予我认为的像我一样很有可能正确的所有人的所有各种价值观或者理想以同等权重。

如果我们假定我们的价值观或者理想可能或多或少地得到正当性的证明，那么我们应当如何对这个怀疑论的论证作出反应就成了一个令人迷惑不解的问题。假定那些最好地得到正当性的证明的价值或者理想是我的，这不是既傲慢自大又不合理性吗？此处我可能受到来自两个方向的拉扯。这个假定看起来可能的确有些傲慢自大或者不合理性。但是，主张我千万不能更加珍视我所更加珍视的东西这一点可能是荒谬的。这就像这样一个主张，即，我千万不能把我所信以为真的东西信以为真。而且这个主张可能从根本上削弱了我对这些价值中的任何一个价值的信念。[2]

有来自两个方向的论证。但是没有哪一个论证支持S。再一次地，S占据一个无法辩护的中间地位。如果这个怀疑论的论证成功，那么中立获胜。就像黑尔的"自由主义者"一样，我应当给予每一个消息灵通和合乎理性的人的价值和理想以同等权重。[3]假设这个论证失败。如果我应当给予我的价值和理想以更大的权重，那么，我也应当现在给予我现在所珍视或者相信的东西以更大的权重。对前一个主张的那个论证，当得到贯彻的时候，证明后面一个主张的正当性。

这些论证之间的冲突提供了内格尔所说的主观的与客观的之间的冲突的一个例证。[4]那些客观性的主张在此是那些具有主观际性（intersubjectivity）的主张。它们把我带到我自己的生活疆界之外，而且告诫我给予他人的价值或者理想以权重。而那些具有主观性的主张则是我现在所相信的主张。它们是我目前观点的那些主张。当我们关心价值或者理想的时候，我们不能够可辩护地主张说具有重要意义的单位是一个人的整个生活。我们不能够主张说，我现在应当给予而且只给予我在任何时候要么有过、要么有、要么将有的所有价值或者理想以同等权重。再一次地，我们不能够既要求时间中性，又拒斥人际中立的要求。刚刚所描述的论证中两者算来都不利于自利论。

61.纯粹的过去愿望

这些主张只适用于其丧失涉及主意的改变的那些愿望。就无数的愿望而言这一点并不为真。即便像一些人所主张的，所有的愿望都涉及一种价值评估，也是如此。值得关心的对象和可能的对象的范围很广。鉴于我们的主意和生命的限度，我们每人只能够强烈地关心这些对象中的一些而已。而且在不必认为我们现在所关心的对象更值得关心的情况下，我们的关心也可能从其中的一个转移到另一个。在上面那个我于火车旅行的终点强烈希望那个陌生人取得成功的例子当中这一点是很清楚的。我以后将丧失这个愿望，但这将不是因为我断定这个陌生人是否成功不再那么重要。

再举一个事例。年轻的时候我最想成为一名诗人。这个愿望在其持续性方面并不是受条件限制的。只有当这以后仍然是我所想要的东西时，我才不想要成为一名诗人。既然我老了，我已经失去了这个愿望。在我已经改变了我的意向在这较受限定的意义上我已经改变了我的主意。但是我并未断定诗歌不再那么重要或者不再那么值得。尽管我现在并没有要写诗的愿望，但是我过去的愿望给予我一个现在力图去创作诗歌的理由吗？

既然我这个愿望的丧失并不涉及我的价值判断方面的改变，那么自利论者也就只有下述这些可供选择的余地。他会保留这样一个主张，那就是，一个合理的行动者应当是时间中性的：他应当力求做那将最好地实现其整个一生中的所有愿望的事。那么他必须接受的是，根据他的理论一个合理的行动者应该赋予他过去的愿望中的那些其丧失并不涉及价值判断方面的改变的愿望以权重。惟一的一些例外是其持续性方面有条件的那些过去的愿望。因此，他必须主张，我有强烈的理由现在努力创作诗歌，因为这是我这么多年来最想要做的事。我具有强烈的努力创作诗歌的理由，尽管我不再有这么去做的丝毫愿望。我们绝大多数人会发现这个主张难以相信。

如果自利论者同意，那么他必须拒斥时间中立的要求。他必须主张说，不给予一个人过去的愿望以什么权重并不是非理性的。但是一旦他放弃时间中立，或者一旦他放弃有关何时我们具有某个愿望不可能在合理性上具有重要意义这个主张，那么，他的理论的其余部分就更加难以辩护了。他必须仍然坚持在我们当前的与我们未来的所有愿望之间的时间中立。他必须坚持认为我应当给予我可预见的将想要的任何东西以同等的权重，即便我现在还并不想要。如果我曾有的那些愿望的确不能够被给予什么权重，这一点就更加难以辩护。如果过去的那些愿望能够具有合理性方面的重要性，那么并不是当前的愿望的那些愿望为什么不能够具有合理性方面的重要性？为什么当前应当被当作宛如是未来的部分来看待？如果不能给予我曾经欲求的东西以任何权重是因为我现在并不欲求，那么，当我现在还并不欲求的时候，为什么我必须给予我将欲求的东西以同等权重？

在前面三节中，我讨论了假定了有关自利的愿望实现论的某个版本的那些自利论者。我对两个结论作了辩护。我们的一些愿望或基于价值判断，或基于理想，抑或基于道德信念。就这些愿望的事例而言，我们必须接受P而不是S。就我们的其他愿望的事例而言，一个S论者有两个可供的选择。他可能坚持认为我们应当力求实现我们的无条件的过去愿望：即那些我们的确曾有但是不再具有的愿望。这是难以置信的。S论者的另一个选择是放弃何时具有某个愿望并不会带来什么差异这样一个主张。如果他放弃这个主张，S论者需要一种全新的解释，以解释为什么我们应当给予我们当前的愿望与我们未来的愿望以相同的权重。这种全新的解释可能难以找到。

如果S论者放弃对时间中立的诉求，那么他必须放弃我所称的他的第二回应。按照他的新观点，我过去的确曾有力求实现我过去的愿望的理由；但是，因为我不再具有这些愿望，我并不具有什么现在力求实现它们的理由。如果是这样，那么他必须舍弃他有关任何理由的力量都是历时地展开的主张。

我的另一个结论也从根本上削弱了这一主张。我具有我当前的价值、理想抑或道德信念所提供的行动理由。如果我后来改变主意，那么我将具有相反的行动理由。就这些理由的事例而言，我们必须接受P而不是S。这些理由的力量并不是历时地延展的。

62.较少关心一个人的较远的未来是非理性的吗?

S论者未能够主张说任何理由的力量是随时间而展开的。他因而必须诉诸他的另一个论证。他必须主张说，那个对自身的偏向是至高无上地合乎理性的。我现在将把这个偏向与另一个常见的关心模式加以比较：较少关心一个人的较远未来。既然这是S论者们最喜欢的对象，那么我将在他们认为最令人信服之处对他们的理论提出挑战。

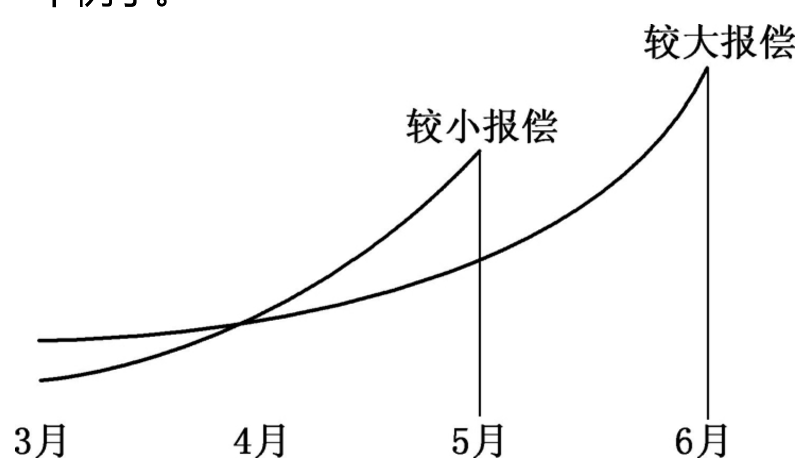
我也将从那个对于愿望实现论而言特有的东西转到那个对于有关自利的所有令人信服的理论而言共同的东西。根据所有这些理论，快乐论至少是真理的一部分。享受、幸福与趋乐避苦是那些使我们的生活变得更好的东西中的一部分。这些在那些我将要予以讨论的事例中是重要的。我之所以选择这些事例部分是因为它们是简单的，部分是因为它们在许多人看起来是自利论最有说服力的那些事例。

边沁主张，在决定任何未来快乐之价值的时候，我们应当考虑我们将多快享受到它。[5]C·I·刘易斯提出，这可能一直在模糊地提示着"较近的快乐一般说来更加明确。"[6]但该主张可能会是多余的主张，因为边沁告诉我们要直接考虑未来快乐的可能性。如果我们严格地对待他的主张，那么它告诉我们优先选择较近的快乐只是因为它们更近一些。它使边沁专心致志于这样的观点，那就是，"尽管我们应当合理地关心未来，但是我们应当因为它更遥远而较少予以关心——而且这完全不取决于任何对更加遥远的怀疑"。刘易斯称之为"零碎的谨慎原则"。正如他所承认的，"它表达的是人们的确倾向于采取的一种态度。"但是他认为该原则的非理性之处太明显，以至于不值得予以讨论。

我称这一态度为近期偏向。休谟描述过这个倾向的表现方式之一：“在思考12个月以后我将要作的一种行动时，我总是决意选择那个较大的善，无论到了那时它将是较近的、还是较遥远的……但当我在较为接近的时候……对于当前的善产生了一种新的倾向，使我难以不变地坚持我的最初目的和决心。”[7]

休谟的话所提示的是，这个偏向只适用于最近的未来。但是一种更为精确的表述是这样的。我们具有一种相对于时间的贴现率，而且我们对于越近的未来给予的贴现率越大。这就是我们为什么并不“不变地坚持”我们的“决心”的原因。这里有两个例子。我决定，当我5分钟之后从我的腿上除去石膏的时候，我将一下子把它除去，现在我宁愿选择片刻的极痛，而不选择把石膏从一根接一根的汗毛上剥离下来的这一漫长的不舒服。但是当那一刻到来的时候，我却改变了我的决定。同样，我决定5年后开始我的职业，我将把我职业的前半段时间用在某个单调乏味但很有可能会在后半段把我推往最高位置的一个岗位。但是当那一刻到来的时候，我再次改变了我的决定。在这两个事例中，从远处看来，某个糟的事情因其随后的善的缘故好像值得承受。但是，当糟和善这两者都更近了的时候，天平则向另一边倾斜。

下面是另一个例子。



每条弧线的高度表明我在任何一个时刻对两个可能的未来报偿之一的关心程度。我对较为遥远的未来的关心较少；而且我的关心中所

减少的量值呈现出在越近的未来则越大。这些弧线恰在我得到这些报偿之前最为陡峭这个事实就表明了这一点。对这些主张重新加以描述可能有所帮助。如果两个相似的事件之一将晚一个月发生，我现在对此则关心得更少。我的关心将与我对早一个月的另一个事件的关心成比例。当这两个事件在未来中更遥远的时候，我当前对它们的关心中成比例的差异将更少。当早一些的事件就处于当下的未来的时候，那么成比例的差异将是最大的。

这些主张解释了为什么在我的图表当中两条弧线是相交的。当它们相交的时候，我的优先选择发生了变化。从3月来判断，我宁取6月更大的报偿而不是5月较小的报偿。从4月末来判断，我则宁取5月较小的报偿。[8]

接下来考虑一下某个具有不同折扣率——指数贴现率——的人。这样的一个人按照每月百分之n的一个常量给未来贴现。这个人对两个未来事件的关心程度之间总是存在同一成比例差异。他的优先选择因而并不以上述方式发生变化。正因为如此，他可能错误地认为他没有什么贴现率。[9]

近期偏向通常以更为简单的方式表现出来。当筹划未来的时候，我们通常把快乐纳入较近的未来而把痛苦往后推迟。但是这个偏向通常会被另一种对待时间的态度所掩盖。后者乃是未来偏向。这个态度并不适用于那些令我们自豪或惭愧的事件：即那些给我们的生活画面添光加彩或抹黑的事件。就像近期偏向一样，这个未来偏向最为明显地适用于它们本身令人惬意或痛苦的事件。当这样的一些事件处于未来而不是过去的时候，对它们的思考给我们造成的影响更大。企盼一种快乐一般来说要比回顾一种快乐更为令人惬意。而且就痛苦的情形而言，前后之间的差异甚至更大。比较一下一个英国小学生被打前后的心态。

我们的行动方式看起来时常表明我们并非近期偏向：我们把痛苦纳入较近的未来而且把快乐往后推迟。未来偏向为此提供了解释。我

们想把痛苦抛在我们后面而把快乐保持在我们前面。既然第二个偏向与第一个偏向相冲突，我们以这两种方式行动的倾向不能够表明我们对近期没有什么偏向。我们的近期偏向可能总是被未来偏向所压倒。我仍然记得，在吹熄了我10周岁生日蛋糕上的蜡烛之后，我决心在未来总要最后吃而不是最先吃最好的部分（蛋白杏仁糖果）。

再看一个例子。假设我必须选择什么时候接受某个痛苦的治疗过程。如果我等上一年，那时医院将添加新设备，治疗将只有一半的痛苦。而且假设我的贴现率一年之内将跌到一半。如果我推迟治疗一年，那么我现在对它的关心就只有四分之一。它本身将是一半的痛苦，而且我现在对它的贴现率是一半。但是如果我推迟该项治疗，我将具有一整年的痛苦预期。这样一种前景，即便已经经过贴现，对我现在而言也可能看起来比那个即刻治疗的前景更糟。如果是这样，尽管我存在近期偏向，我将选择将有两倍痛苦的现在予以治疗。

有一些人并不是更加关心近期的东西。一些人甚至更加关心长远的东西。节约或推迟满足的习性能够是强迫性的。但是我们在此无需判定究竟有多少人具有这种近期偏向。

另一个问题更为重要。人们时有这样的主张，那就是，这个偏向总是由想象力的某种失败或者某种虚假的信念所导致。例如，有主张说，当我们想象更遥远的未来中的痛苦的时候，我们对它们的想象较不那么栩栩如生，或者错乱地认为它们由于某种原因将较不实在或较少充满痛苦。自从柏拉图作此主张以来，[10]就一直经常为人所重复。于是皮古主张说我们有“望远功能方面的缺陷”。[11]而且这一主张深深地嵌入了我们的语言之中。一个人如果较不关心他自己的更遥远的未来，则是不谨慎的或无远虑的——这些词在拉丁语中的意思是说一个人没有对未来的远见。而一些经济学家则称这一态度为近视症。

柏拉图的主张通常正确。就许多人的情形而言，它为他们偏向于近期提供了部分解释。如果这个主张总是为真而且提供整个说明的话，那么这一点会是相当重要的。如果是这样的话，那么这个偏向则

永远不会逃过理想的审慎过程。这会减少S与某些P版本之间的冲突程度。但是正如我在第73节和附录C中所论证的，S和P仍然不会相契合。

尽管这是一个事实问题，但是我相当肯定的是，柏拉图的主张通常是虚假的。一种检验方法会是这样的。在一个实验中，某人必须就其是否为了某种快乐之故而忍受某种痛苦作出决定。这个人知道，一旦他作出这个决定，他将吞服诱发他忘记这个决定的一种药丸。这使快乐或痛苦的预想成为不相关的。这个人还知道，直到他作出决定的那一刻之前，我们才告诉他这个痛苦和快乐的时间选择（timing）。我们仔细地描述了两种经验会包含的东西。所以这个人可以完全知情地作出决定，他可以尽其所能，栩栩如生地想象忍受这个痛苦和享受这个快乐会是什么样子。然后我们告诉他说，这一痛苦会是即刻到来的，而快乐将会延期一年。现在这一快乐在他看来并不栩栩如生吗？至少就我自己的情形而言，我肯定它不会如此。假设，如果这一个痛苦会是即刻到来的而快乐将会延期一年，那么这个人具有不选择两者中的任何一个的温和倾向。他断定这个快乐不足以让他值得忍受这个痛苦。然后我们又告诉他我们搞错了：这个快乐会是即刻到来的，而这个痛苦将延期一年。我认为很有可能的是，这个人的优先选择现在会出现变化。他现在可能断定，为了享受这个快乐之故，这个痛苦是值得忍受的。既然，当这个人设想这两种经验的时候并不知道它们的时间安排，那么他的优先选择方面的这样一种变化不会产生自这样一个所宣称的事实，那就是，后来的经验在想象中总是看起来并不栩栩如生。我们就会有充分的理由相信，这个人是基于近期偏向，而且在某种意义上逃过理想的审慎这一劫。

我认为这种近期偏向是普遍的。但是为了避免争论，我们可以讨论一个想象的人。这个人之所以更关心他的较近期的未来，只是因为它更近一些而已；而且即便当他知道有关事实、思虑清楚的时候也是

这么做。我将称这人为重近者。正如我所认为的，至于实际上是否存在许多像这个人一样的人，这一点并不影响论证。

既因为相关事实是值得怀疑的，又因为有关自我利益的敌对理论之间的分歧，所以，对某人而言什么会是最好的或者什么是最符合利益的这一点时常并不清楚。但是就所有令人信服的理论而言，有一点是得到认同的。当我们在决定什么符合一个人的利益的时候，我们应当因不确定性而打折扣，而不应当仅仅因遥远性而打折扣。我们不应当给予这个人更遥远的未来以较小的权重，也不应当给予他目前的愿望以更大的权重。

我们可能认为，我们有时应当给予某个他人的当前愿望以更大的权重。我们可能认为，压抑这个人的当前愿望、强迫他去做对他而言将会最好的事这一点是错误的。但我们会因为它侵犯了这个人的自主权而认为这一点是错误的。这符合我的这样一个主张，那就是，按照有关自我利益的所有令人信服的理论，在决定什么对某人而言会最好的时候，我们不应当给予这个人当前的愿望以更大的权重。我们应当给予这个人的生活的所有部分以平等的权重。

按照自利论，当某个人做他明知对他自己来说更糟的事的时候，他的行动是非理性的。我所说的那个想象出来的人时常以这种方式行动。因为他是基于近期偏向，重近者经常审慎地推迟那些痛苦，以预见到它们变得更糟为代价。在这些情况下，他是在做他明知对他自己来说更糟的事。

现在让我们把他的态度与一个利己的人的态度相比较。近期偏向比那个有利于自身的偏向较少合乎理性吗？我们的肯定回答对于S的辩护而言是相当重要的。

重近者知道相关的事实，而且是在清晰地思考。我们应当再添加一个假设。那些具有某个偏向的人们可能希望不带有这个偏向。这一点就对近期偏向这一情形而言是相当普遍的。在描述了这个偏向如何使他的行动有悖于自己的利益之后，休谟写道：“对于这个自然的弱点

我可能非常抱憾”。[12]我们应当假定重近者没有这样的遗憾。只有这个假定才使我们所做的比较显得公正，因为那些利己的人们被典型地假定为对他们那些有利于自身的偏向并不感到遗憾。况且，在拒斥S方面，重近者将诉诸P；而P的那个批判性的版本可能对行动者希望自己所不具有的那些愿望大打折扣。该版本可能主张，我们应当给予这样的一些愿望较少的权重，甚至不给予它们什么权重。我们因而应当假定重近者并不为他的近期偏向而感到遗憾。

它可能遭遇到这样的诘难，那就是，当他身受这个偏向的后果之苦的时候，他必定为具有该偏向而感到后悔。但是正如我在第71节中所解释的，重近者既从不对他现在具有这个偏向而遗憾，也从不对他在较近的过去具有过这个偏向而感到遗憾。他充其量会为曾在更远一些的过去具有过这个偏向而感到遗憾。因为他当前的偏向，他就像他所做的那样行动；而且，他从不为这个偏向而感到遗憾。

63.一个自杀性的论证

自利论者应当如何批判重近者？如果让他们选择一个很快到来的轻微痛苦或者后来一些的一个更严重的痛苦，那么他们通常审慎地选择那个更严重的痛苦。而且他时常是优先选择一个很快到来的小快乐而不是一个后来到来的大得多的快乐。因而他必定附和休谟主张说：“我如果选择我所认为较小的福利而舍去较大的福利……那也同样不是违反理性”。[13]这一点——审慎地选择他承认对自己更糟的事——看起来是可能有的最为清楚的非理性事例。S论者或许说：“合理性的第一条规则就是拒斥你明知更糟的事。”

重近者会回答说：“如果两个痛苦之间的惟一区别是一个会更糟的话，我接受你的规则。但是，在我们正在讨论的那些事例中，还有另一种区别。当我在两个痛苦之间进行选择的时候，我既考虑它们会多么痛苦，又考虑我应当多久不得不经受它们。我并非只是在选择我明

知更糟的事。只有当更糟的程度对我现在来说被未来更进一步糟的程度所超过的时候，我才选择了两个痛苦中更糟的那一个。”

该S论者必定回应说，把近这一点考虑进去是非理性的。他或许主张，引用罗尔斯的话来说就是，“纯粹的时间上的位置或与当前的距离，并不是偏爱一个时刻而非另一个时刻的一个理由。”[14]更宽泛一些说，该S论者可能恢复那个时间中性的要求。

我在前面论述过，如果S论者假定有关自我利益的愿望实现论，他就应当舍弃那个时间中性的要求。按照这个有关自我利益的理论，这个要求并不令人信服地蕴含着这样一点，那就是，我们现在应当努力实现我们一度拥有的某些愿望，即便我们现在并不拥有而且后来也将永远不会有这些愿望。在下文中我将假定S论者拒斥愿望实现论，要么接受快乐论，要么接受目标列表理论。根据这个假定，如果S论者要求时间中性，那么他就无需主张我们应当努力实现这样一些过去的愿望。他的主张能够是这样的，那就是，当我们在考虑快乐与痛苦的时候，或者幸福与受苦的时候，纯然的时间选择上的差异不能够具有合理性方面的重要性。

S论者应当如何支持这个主张？为什么不应当把时间考虑进去？他或许说：

> 一个某事何时发生方面的纯粹差异不是其性质方面的一种差异。一个痛苦处于更遥远的未来的这个事实，将不会使之到来的时候变得较少充满痛苦。

这是一个杰出的论证。这是迄今对重近期偏向的一个最好的诘难。但是S论者不能够使用这个论证。它是一把双刃剑。同一论证能够用来针对自利论。正如重近者把何时感受到一个痛苦考虑进去一样，S论者把何人将感受到痛苦考虑了进去。而且一个在何人感受到痛苦方面的纯粹区别并不是其性质方面的一个差异。一个痛苦是某个他人的痛苦这个事实并不使之变得较少充满痛苦。

S论者考虑进去（1）那些痛苦会如何糟糕，以及（2）何人感受到它们。他因而有时候选择两个痛苦中更糟的那一个。他有时候选择一个对某个别人而言更糟的痛苦，而不是对他自己而言较轻的痛苦。（可能看起来他总是会作出这种选择。但是这假定了S论者必定是纯粹自私的。正如我所主张的，这是一个错误。一个接受S的人可能爱某些特定的他人。因而，如果他逃避了某个较轻的痛苦，而代价却是施加给他所爱的某个人以更糟的痛苦，那么对他而言可能更糟。）

重近者考虑进去（1）那些痛苦会如何糟糕，（2）何人感受到它们，以及（3）何时感受到它们。他能够对S论者说："如果你能够把何人将感受到某个痛苦考虑进去，我为什么就不能够把何时某个痛苦被感受到考虑进去？"对这个问题可能有一些答案。可能有一些论证表明，个人同一性方面的那些差异具有时间选择方面的那些差异所不具有的一种重要性。我迄今所表明的要点只在于此。在解释为什么时间不能够具有合理性方面的重要性的时候，S论者不能够使用那个明显的、最佳的论证。他不能够诉诸这样一个事实，那就是，一个痛苦并不因为它是较不近的而较少充满痛苦。一个痛苦并不是因为它属于某个别人而较少充满痛苦。

S论者或许说：

> 你误解了我的论证。一个痛苦处于你的更遥远的未来这一点不能使之对你变得较少充满痛苦。但是，一个痛苦是某个别人的痛苦这一点却使之对你变得较少充满痛苦。如果它是别人的痛苦，它将一点都伤不到你。

这些话中的第二句提出了一对主张。一个痛苦处于我的更遥远的未来这一点并不使之变得要么（a）较少充满痛苦，要么（b）较不属于我。（a）为真，但是并不相关，因为它所诉诸的那个诘难同等程度地适用于自利论。一个痛苦是某个别人的痛苦这一点并不使之变得较少充满痛苦。（b）同样为真。一个痛苦处于我的更遥远的未来这一事实并不使之变得较不属于我的痛苦。但此点为真并不是一个论证。

S论者在抨击重近者时所需要主张的是：何人感受到痛苦方面的区别具有极大的合理性方面的重要性，而何时一个痛苦被感受到方面的差异则并不会具有合理性方面的重要性。（b）所指出的一切是，这是一些不同的差异。时间并不像个人同一性一样。凭它自身，这不能够表明时间较不重要。

我现在将总结一下这些主张。S论者必须批判重近者。根据S，我们既应当把痛苦程度方面的差异考虑进去，又应当把受苦者同一性方面的差异考虑进去。重近者还把时间选择方面的差异考虑了进去。S论者尚未说明，个人同一性方面的差异具有时间选择性方面的差异所缺乏的合理性方面的重要性。对这个主张可能有一些论证。但是我还没有给出这样的一个论证。S论者不会使用那个最佳论证。他不会以主张时间选择方面的那些差异不是痛苦程度方面的差异来拒斥时间选择方面的那些差异。它们也不是个人同一性方面的差异。S论者也不会基于时间选择性方面的差异不是个人同一性方面的差异来拒斥时间选择方面的差异。这些是不同的差异这一点并不能表明一个具有别人所不具有的合理性方面的重要性。

64.过去的或者未来的受苦

S论者或许主张，没有什么论辩的必要。我们不会论证每一件事情；有一些事情就得要假定。而且有关他当前的主张他就可能这么说。他可能说，当我们把"它发生在何人身上？"与"它何时发生"这些问题加以比较的时候，我们清楚地看到，只有前一个问题具有合理性方面的重要性。我们清楚地看到，如果某个痛苦将被某个别的人所感受到，那么对它较少关心就并非是非理性的；但是，只是因为某个痛苦被一个人本身何时感受到方面的差异而对它较少关心则是非理性的。

情况是这样吗？对近期偏向并不是我们的相对于时间的惟一偏向。我们还偏向未来。这个态度是非理性的吗？

考虑一下我的过去的手术或未来的手术。

事例一。我在某个医院，要做某种手术。因为这是一种完全安全的手术，而且总是取得成功，我对于手术结果没有什么害怕之处。外科手术可能是短暂的，也可能是漫长的。因为我得同医生配合，所以我不能够用麻药。以前我曾经做了一次手术，而且能够记得如何充满痛苦。在一项新方法下，鉴于手术如此充满痛苦，故现在事后都使病人忘记手术。某种药物清除掉他们最后数小时的记忆。

我刚刚醒过来。不能够记起入睡过。我询问我的护士是否已经决定何时给我动手术，手术得要多长时间。她说她知道有关我与另一个病人的事实，但是她记不清哪些事实适用于哪一个人了。她只能够告诉我下述情况为真。我可能是昨天动了手术的那位病人。在这种情形下，我的手术是迄今所实施的最漫长的手术，持续了10个小时。我也可能是有待今天晚些时候进行一次短暂手术的那位病人。要么我的确遭受了10个小时的手术为真，要么我将遭受1个小时的手术为真。

我请求那位护士弄清楚哪一个为真。当她离开期间，对我而言我优先选择哪一个为真是清楚的。如果我得悉第一种情况为真，我将大大地松一口气。

我的未来偏向使我在此对于我的痛苦是在过去这一点感到如释重负。我的近期偏向可能同样使我对某个痛苦已经被延迟这一点感到如释重负。无论在哪一种情形下，我或许都会为我的磨难优先选择某个与之不同的时机，即便由于这个不同的时间该磨难会糟糕得多。较之今天晚些时候1个小时的痛苦，我或许像重近者一样优先选择明年10个小时的痛苦。或者，就像在这个事例中的情形一样，我或许优先选择昨天10个小时的痛苦。

这第二个优先选择是非理性的吗？我应该代之希望我是痛苦尚待到来的第二个病人吗？在我讨论这个问题之前，我应当解释一下该事例的一个特征：诱发性失忆。

一些学者主张，如果我的未来的某个部分将并不借助记忆与我的生活的其余部分相连，那么我就能够合乎理性地忽视在这个期间将发生在我身上的事。对这些学者而言，双重剂量的失忆就像麻药一样好。如果我在受苦的时候将没有什么记忆，而且我后来将没有什么对于我受苦的记忆，我则无需——他们主张——关心这个未来的受苦。这是一个有争议的主张。但是，即便它得到正当性的证明，它也并不

适用于我所举的例子。这个例子并不包含双重剂量的失忆。在我充满痛苦的手术期间，我将具有我的所有记忆。固然事后我将被诱发忘记该手术，但是这并不消除我关心我的未来所受之苦的理由。如果我们否认此点的话，那么我们就会不得不主张，当某人已知他将要死去，他得知他将痛苦地死去这个额外事实的时候，他不应当在乎。他后来将不记得这些痛苦。

如果我们设想自己处于那个今天晚些时候将受1小时苦的病人地位，那么我们绝大多数人会有所关心。即便知道我们以后将不会记得这1小时的痛苦，我们也会有所关心。而且现在我可以解释为什么我所举的那个事例包含诱发性的失忆。这提供给我们那个正确的比较。如果我已经获悉我是第二个病人，那么我就处于下面的这个心态。我相信我今天晚些时候将遭受1小时的痛苦，而且我能够大致想象这个痛苦将有多么可怕。这足以令我关心。如果相反我获悉我是第一个病人，那么我就处于那个全然可比较的心态。我相信昨天自己的确经受了10个小时的痛苦，而且我能够大致想象那场痛苦会有多么可怕。我的心态只是在我正在讨论的两个方面有所不同。我的信念具有一个不同的时态，它是有关过去而不是有关未来。而且它是一个有关10个小时的痛苦而不是1个小时的痛苦的信念。如果我并非只是相信我昨天遭受了痛苦，而且还记得这场痛苦，就会使那个比较出现混乱。当我相信今天晚些时候我将受苦的时候，我没有什么与这个未来所受之苦的记忆可比的。而且痛苦的记忆是多种多样的；一些本身充满痛苦，而另一些则不然。假如有关我过去的痛苦我只会具有我有关未来的痛苦所会具有的——伴有对该痛苦之可怕的想象能力的一个信念的话，那么它就因而从该事例中除去一个不相干的且令情况更加复杂的特征。

诱发性失忆净化了该事例。但是它仍然可能引起怀疑。我因而补充

事例二。当我醒来的时候，我的确记得昨天的一段长时间的受苦。但是我无法记得这段时间有多么长。我问护士我的手术是否完成，或者是否需要进一步的外科手术。如前所述，她知道两个病人的有关事实，但是她不能够记起我是其中的哪一个了。如果我是第一个病人，那么我昨天就经受了5个小时的痛苦，而且我的手术结束了。如果我是第二个病人，那么我昨天经受了2个小时的痛苦，而且我将在今天晚些时候遭受另一个小时的痛苦。[15]

在事例二中，没有失忆症；但是这并不带来什么差异。要么我遭受了5个小时的痛苦，而且再没有更多的痛苦要到来；要么我遭受了2个小时的痛苦，而且还有1个小时的痛苦要到来。我会再次宁愿第一个情形为真。我会宁愿我的生活包含更多小时的痛苦，如果这意味着之后不再有这一痛苦到来的话。

如果我们设想我们自己处于我在这两个事例中的位置，我们绝大多数人会具有我的优先选择。如果我们并不知道我们是已经遭受了数小时的痛苦，还是将在后来遭受1个小时的痛苦的话，我们绝大多数人会强烈地宁愿第一种情况为真。如果我们能够使之为真的话，我们会毫无疑问地这样做的。如果我们信教的话，我们可能会祈祷这一点为真。按照一些说法，这是一个可以想象到的对过去产生影响的方法。上帝之所以可能已经使某个过去的事件发生，只是因为在那时他预见到了我们后来的回望祈祷，而且他选择了恩准这个祈祷。即便我们并不相信我们能够以这种方式，通过上帝的神恩，造成我们的痛苦是在过去，我们也会强烈地宁愿它是在过去，就算代价是它持续十倍长也在所不惜。

这种优先选择是非理性的吗？我们绝大多数人会给出否定的答案。如果S论者接受这个答案，那么他必须放弃他有关"何时？"这一问题在合理性上并不重要的主张。他不会主张说，痛苦的时间选择上的一个纯粹的差异，或者它与当前时刻的关系上的一个差异，"本身不是对它予以或多或少关注的一个合理的根据"[16]。一个痛苦是在过去还是在未来，纯粹是它与当前时刻的关系中的一个差异。而且，如果更多关注那些处于未来的痛苦并不是非理性的，那么何以更多关注处于

更近一些的未来中的痛苦就是非理性的？如果S论者承认对时间上中立的一个背离是可以得到辩护的，那么他如何能够批判另一个？

65.因果关系的方向

S论者或许说：“既然我们不能够影响过去，那么这就是对之较少关心的一个良好根据。不存在什么对于近期偏向的这样一种正当证明。”

这一点能够得到回答。我们首先能够指出，即便是当未来就像过去一样不能够被影响的时候，我们仍然偏向未来。假设我们身陷囹圄，今天晚些时候将遭受折磨。在这样的一些情形中，当我们相信我们未来的所受之苦不可避免的时候，我们对它的态度并不落入我们对待过去的所受之苦的窠臼。我们不会认为，“既然这场折磨是不可避免的，那么就等同它已经是处于过去。”当这样的不可避免的痛苦处于过去的时候，我们则大大地松一口气。在这样的一些情形中，未来偏向不能靠诉诸因果关系而得到正当证明。与过去的痛苦不同，我们关心这样的未来痛苦的根据并非我们能够影响它们。我们明白，我们不能够影响它们。我们之所以关心这些痛苦，只是因为它们尚未处于过去。

S论者可能回应说：“这样一种正当证明无需在每种情形中都成立。当我们是在讨论一个一般性的态度的时候，我们必须满足于一个一般性的真理。这样的一些态度不能够是‘微调’。在绝大多数情况下，事件是否处于过去，对应于我们是否能够影响它们。这足以给予未来偏向提供正当证明。如果我们缺乏这个偏向的话，我们就会与之同等程度地去关心我们所不能够影响的过去的痛苦与快乐了。这就会分散我们对于我们所能够影响的未来的痛苦与快乐的注意力。因为我们的注意力会被这样分散，所以我们在努力获得未来的快乐和避免未来的痛苦方面的企图就会不那么成功。这对我们而言会更糟糕。”

我们会回答说："如果这是真的，那么还有另一个类似的真理。如果我们同等地关心我们更远的未来中的痛苦与快乐的话，那么这就会分散我们对更加近期的未来中的痛苦与快乐的注意力。如果我们想要减轻我们未来的所受之苦，我们就应当更多地注意较近的未来中那些可能的痛苦与快乐，因为我们用以避免或者减轻这些痛苦的时间更少。相似的主张适用于未来的快乐。我们对较近的未来加以影响的需要更加紧迫。如果你的主张证明对于未来之偏向是正当的，那么这个主张则证明对于近期偏向是正当的。"

我们会补充说："甚至在我们不能够影响到近期的未来的那些特殊情况下，我们对它也更加关心。但是这些情形对应于我们其中不能够对未来施加影响的那些特殊情形。这两个对待时间的态度都大致对应于这些有关因果关系的事实。因而你的主张不能够说明这些态度中只有一个是可以得到辩护的。"

S论者可能说："你忽略了一种差异。我们能够直接按照对于近期偏向而行动。如果我们预定今天晚些时候要遭受1个小时的痛苦，我们可能能够以使之更糟为代价来推迟这个痛苦。我们可能像重近者一样，把这个痛苦换成明年10个小时的痛苦。但是我们不会把这个痛苦换成昨天10个小时的痛苦。我们不会以使那些未来的痛苦更糟为代价来把它们置于过去。该重要差异就在于此。既然我们既能够影响近期的未来，又能够影响遥远的未来，那么我们的近期偏向时常造成我们的行动有悖于我们自己的利益。这个偏向对于我们来说是糟糕的。相反，由于我们不能够影响过去，所以，我们的未来偏向决不会造成我们的行动有悖于我们的利益。这第二个偏向对我们而言不是糟糕的。此乃只有这第二个偏向可以得到辩护的原因。"

对于这一说法有三个回应：（1）这个论证具有一个虚假的前提。一个态度对我们而言是糟糕的这个事实并不表明它是非理性的。它充其量只能够表明我们应当努力改变这个态度。如果某个我所爱的人被杀，在经过一段时间之后我应当努力节哀。但是这并不表明我没

有理由哀伤。哀伤仅因带来不快乐并非就是非理性的。对于"你的悲伤是徒劳的"这个主张，休谟回应说，"的确如此，而这正是我感到难过的理由"。[17]

（2）即使（1）遭到否定，这个论证也失败。它假定，重要的是某个事物对我们而言是否糟糕。这是以尚待证明的假定为论据。S论者一直是在谴责近期偏向。如果我们有这个偏向，我们则对我们较近的未来关心得更多一些；而总体上说来对我们将更糟糕的东西在较近的未来对我们则可能更好。如果我们的偏向是可以辩护的，那么我们因而能够否定重要的是某个事物对我们而言是否糟糕这样一个假设。既然，如果我们的偏向是可辩护的，这个假设则能够被否定，那么它就不能够有助于表明我们的偏向不是可辩护的。

（3）尚未表明近期偏向对我们而言是糟糕的。因为我们有对较近未来进行影响的更紧迫的需要，所以近期偏向在某些方面对我们而言是好的。但是，且让我们假设这个偏向总的说来对我们而言是糟糕的。未来偏向也是如此。正如我后面将予以解释的，如果我们不是更加关心未来的话，对我们而言会更好。上述那个论证还有另一个虚假前提。未来偏向对我们而言并不糟糕这一点并不为真。

S论者必须谴责近期偏向。如果他的批评诉诸时间中性，那么他必须批评近期偏向。凭借诉诸有关因果关系的那些事实，S论者努力避免这个结论。但是这个尝试已经失败了。

在谴责近期偏向方面，S论者可能说："既然我们对于影响近期的需要更加紧迫，那么近期偏向就完全是自然而然的。进化把这个偏向给予所有动物并不令人诧异。但是，既然我们是理性的，那么我们能够上升到我们从进化所遗传的东西之上，而且能够批判性地对之予以审视。我们能够看清，这个偏向不能够是合乎理性的。某个痛苦处于更近的未来这一点不能够成为对之更加关心的一个理由。时间选择上的一个纯粹差异不能够具有合理性上的重要性。"[18]

如果S论者提出这个主张，那么有关近期偏向他必须提出相似主张。他或许说："既然我们不能够影响过去，那么对之较少关心则是自然而然的。但是这个偏向不能够是合乎理性的。当我们不能够影响近期的时候，这一点最清楚。某个不可避免的痛苦是在未来而不是在过去这一点不能够成为对之更加关心的一个理由。当它处于过去的时候就如释重负是非理性的。"

在我的过去的手术或未来的手术中，我会宁愿我昨天的确遭受了数个小时的痛苦为真，而不是我在今天晚些时候将遭受1个小时的痛苦为真。这不是一个据之我能够作出的优先选择。但是这一点对于这个优先选择是否非理性这个问题并不相关。S论者不能够主张说，因为我不能够据之而行动，所以它并不是非理性的。他会说，"这个优先选择太荒谬了！对于不能够据之而行动，你应当感到感激。"而如果他仍然坚持我们对我们自己的关心应当是时间中性这个主张，这则是他所必须说的。如果他谴责近期偏向是因为当某个痛苦被感受到的时候，未来的偏向不能够具有合理性方面的重要性，那么他必须谴责未来偏向。他必须主张，当某个痛苦处于过去的时候感到宽慰是非理性的。我们绝大多数人会感到这难以置信。如果S论者坚持认为我们应当时间中性，我们绝大多数人会有异议。

66.时间上中立

S论者可能会改变其立场。他可能会谴责近期偏向，但并非基于"何时"这个问题不能够具有合理性上的重要性这个一般根据，而是基于一个更加特定的根据。

他可能会转向另一个极端，主张时间上中立是不可想象的。他可能会主张说，我们缺乏未来偏向是不可想象的。如果这点为真的话，他就会再次只是批评这两个态度中的一个。如果我们缺乏某个态度这一点是不可想象的，那么具有这个态度就不能够是非理性的。但是，与未来偏向不同的是，近期偏向显然是某种我们可能会缺乏的东西。

我们可能会同等关心我们未来的所有部分。一些人就是这样。S论者可能会主张说，这是惟一合理的关心模式。

我们可能缺乏未来偏向这一点是可以想象的吗？我们那些对待过去的态度不会就像我们对待未来的态度。一些情绪或者反应预设了对于因果关系的信念。既然我们不能够影响过去，那么这些情绪和反应就不会是回顾性的。于是我们不会形成一种昨天做了某事的意向，也不会坚定地决意要把留在我们后面的事做得最好。

存在一些本质上瞻望性的、在某种意义上不能够由因果关系的方向所解释的心理状态吗？这是一个大问题，对此我无需给出一个完全的回答。考虑一下我们的未来偏向中所包含的那些最为重要的心理状态就足够了。

其中之一是愿望。我们的某些语言提示着，愿望本质上是瞻望性的。比较一下"明年冬季我真想去威尼斯"与"去年冬季我真想去过了威尼斯"。第二个主张有些含糊不清。

我们的语言在此具有误导性。考虑一下

> 我的时间上中立的愿望。我获悉一位老朋友在某个遥远的乡村正奄奄一息。我们过去是在愤怒中分开的，为此我现在很自责。在获悉我的朋友已奄奄一息之后，我有一种请求她原谅我的强烈愿望。鉴于没有办法用电话同她联系，我所能够做的最好是给她寄一封快信，请求得到原谅，并与她道别。一个星期之后，我并不知道我的朋友是否仍然活着，也不知道她是否收到了我的信。我最强烈的愿望是，她在死前收到了我的信。

如果愿望本质上是瞻望性的，那么我必然被当作处于两个心理状态之中：一个有条件的愿望和一个有条件的希望。我一定被说成是想要我的朋友，如果还活着，在死前收到我的那封信；而且希望，如果她死了，她在死前收到了我的那封信。但是这种表述，即便是语言上所要求的，也是具有误导性的。在此区分出愿望与希望这两个心理状态是把一个本质上单一的状态再分开来。我的"希望"在其本质上和力度上就像我的"愿望"一样。我所想要的是我的朋友先于死亡之前收到

这封信。只要这些事件以这样的顺序发生，至于它们是处于过去还是未来我都无所谓。

因而，即便有点改变概念，最好也要说我们能够有一些有关过去的愿望。这是一种变化这一点并不清楚。例如，我可能想要这样一件事情是真的，就是我昨晚醉酒的时候没有做出什么丢人的事情。而且我可能因其本身之故而想要这是真的，而不是因为它对我的未来的可能影响。同样地，在读过济慈[19]或者凡·高[20]信之后，我可能想使这样一点是真的，那就是，他们明白自己的成就有多么伟大。

在这些事例中，我并不知道真相如何。假设我确实知道我昨晚做出了丢人现眼的事情。我能够想要我并没有丢人现眼这一点是真的吗？称之为但愿会更自然一些。但是这一区分看起来也并不重要。当我获悉我自己丢人现眼了之后，我的愿望——我未丢人现眼——成为一个但愿。但是这个但愿可能绝不比那个愿望弱一些。

如果我们主张我们能够想要某个我们明知虚假的东西为真，那么这可能改变了愿望这个概念。我们无需断定我们概念中的这个变化是否会是一种改进。我们正在讨论的是一个与之不同的问题，即我们是否能够具有一些有关过去的愿望。我已经主张说我们能够，即便我们的某些语言提示着我们并不能够。我们能够用我们语言的其他部分来表达这样的一些愿望。我们能够说，就像我们已经所做的那样，我们想要某个事件确实发生了或者并未发生这一点是真的。

可能有诘难说，愿望本质上是与可能的行动联系在一起的。这就像"应当"蕴含"能够"这一主张一样。按照这个观点，我们不能够具有一些我们会不可能据此行动的愿望。从这个一般性的主张，我们会推演出这样一个特殊的主张，那就是，我们不能够具有一些有关过去的愿望，因为我们不能够影响过去。[21]

这个一般性的主张为假。当然，在愿望和行动之间存在密切的联系。如果我们强烈地想要某事为真，我们就应当努力弄清我们能否使之成真。"想要之原始迹象就是努力去得到"。[22]但是此处愿望居于首

位。在我们能够想要某事为真之前，我们并非一定要明白我们是否会使之为真。

我们能够承认愿望与行动相连的一种方式。如果人们不会行动，那么他们就不会具有愿望。除非我们还具有了行动的概念，否则我们不会具有愿望的概念。但是我们能够具有一个特定的愿望而无需能够按此行动。即便当我们知道不仅我们本就不可能，而且其他任何人本来也不可能使之成真的时候，我们也能够想要某事为真。毕达哥拉斯学派曾想使2的平方根为一个有理数。这个愿望的实现在逻辑上是不可能的。既然我们能够具有一些即便全能的上帝都无法实现的愿望，那么特定的愿望并不与可能的行动相连。这消除了据以否定我们能够具有一些有关过去的愿望的根据。

下面我们能够考虑一下这个讨论中最重要的那些心理状态了：对某个未来事件的期待，以及其反面——充满痛苦的或者令人忧伤的预想。这两个心理状态本质上都是指向未来的。但这可能又是一个表面上的真理。会有指向过去的可比状态吗？

可能有人认为，我们实际上具有这样的一些回顾状态。对于未来的偏向并不适用于许多种类的事件，比如那些令我们自豪或惭愧的事件。尽管认识到过去的一个成就可能给我们带来快乐，但是这与期待并不类同。我们正在讨论的我们的态度不是有关我们的生活包含某些事件这个事实，而是有关我们在其他生活时刻经受这些事件的经验。为简单性起见，我一直在讨论的是对本身纯粹愉快或痛苦的那些经验的态度。我们事实上以我们期待未来快乐的方式回顾过去的快乐吗？

再度存在着一种由记忆所带来的复杂性。这些记忆本身能够是愉快的或是痛苦的。我们可能享受回忆快乐，而不愿回忆痛苦。但是这两者中没有哪一个与预想的痛苦和快乐严格地具有类比性。我们因而需要考虑我们对待这样的一些过去的痛苦与快乐的态度，那就是，我们知道它们，对之却并没有痛苦或愉快的记忆。

考虑一下我过去的那些煎熬。

事例一。我异常健忘。当被问及"你能记得10年前的5月份在你身上发生了什么？"这个问题的时候，我发现有关那个月的事情我什么都记不得了。然后我被告知说，在那个月之初我被发现患上了某种疾病，需要即刻进行4周非常痛苦的治疗。鉴于这个治疗完全取得成功，我没有什么根据去担心未来。当人们提醒我这个事实的时候，它唤起了一种本身并不痛苦的依稀记忆。

令我感到惊讶的是，人们提醒我说，10年前我经受了一个月的极度痛苦。现在我只有对这个事实的依稀记忆，以及想象我所经受的极度痛苦有多么糟的能力。当人们提醒我这个过去的煎熬的时候，我会心烦意乱吗？我会有某种相当于对痛苦进行预想的东西吗？我不会。我对这样的提醒者的反应会是完全漠然置之。

如果我获悉，距今10年之后我将经受一个月的极度痛苦，我就不会漠然置之。我会感到苦恼。但是，如果有人提醒我说，10年之前我经受了这样的一个月，我就绝不会感到苦恼。

既然我们有近期偏向，那么考虑一下事例二可能有所帮助

事例二。在我信以为是5月1日的那一天，我醒了过来。其实那是6月1日。我刚经历了相似的、非常痛苦的一个月，但是得到了完全成功的治疗。为了我对此不会具有痛苦的记忆，由此我忘记了这整整一个月。

我获悉我刚刚经受了一个月的极度痛苦。在此我也不会把这当成坏消息。更确切些说，我会对这样的一个事实感到遗憾，那就是，我一生中的一个月要不得不这样浪费掉。我对所宣称的成功治疗或许有点担忧。而且我也许具有某种担心，那就是，如果所诱发的失忆并不长久的话，我后来将具有对这次治疗的痛苦记忆。但是我一点也不会对在这个月期间我处于极度痛苦之中这个事实感到苦恼。对这个新近的极度痛苦我会完全漠然置之。相反，如果我获悉我将要经受这样的煎熬的话，我就会非常苦恼。

它涉及诱发性失忆这一点可能是对事例二的一个诘难。我因而添加

事例三。在我的实际生活中，我经受严重的痛苦已是家常便饭。我能够记得这些痛苦，但是这些记忆本身并不充满痛苦。我所能够记得的、所受的最厉害之痛苦当属1979年持续了三天的那一次痛苦。

当我忆起这非常痛苦的三天的时候，我一点都不苦恼。在想象的事例一和事例二中，我相信对于我过去的那些煎熬我会漠然置之。在我的实际生活中，我的确对于我过去所受之苦完全漠然置之。

我相信在这个方面绝大多数人像我一样。除非他们的记忆是痛苦的，否则他们对于他们过去所受之苦漠然置之。我知道一些人具有与之不同的反应。这些人主张，即便他们没有什么痛苦的记忆，他们发觉认识到他们过去的痛苦也令人些许苦恼。但是我不知道有什么人具有完全相当于预想之痛苦的东西。

事实上我们对我们过去的痛苦并不持有这个态度。而且我们并不以展望快乐的方式回顾快乐。会有这样的心理状态吗？"回顾"某个过去的事件，除了其时间上的方向性之外，就像展望一样吗？

我们可能说："当我们想到某个未来的事件给予我们快乐的时候，我们展望这个事件。思考过去的事件会带给我们相似的快乐。对于预想的痛苦而言会有相应的回想的痛苦。"

它可能遭受到这样的诘难："你们没有充分陈述展望所涉及到的东西。想到未来的快乐不仅仅固然带给我们快乐，我们还预想这些快乐。同样，我们预想痛苦。预想不能够具有一个回顾性的对应物。"

我们或许回答说："我们可能不能够想象具有这一对应物会是什么样。但是这并不表明它不会被人所具有。那些先天失明的人不能够想象看到是什么样。"

这个回应可能并未答复这个诘难。如果是这样，我们的那些主张可以加以修正。即便回顾不会就像展望一样，但是会令人同等振奋，或者就痛苦的情形而言，令人同等苦恼。这会涉及到我们态度的一种变化。但这种变化是可以想象的。我们能够清楚地描述某个在这方面与我们不同的人。当这样的一个人被人们提醒说他曾经经受了极度痛

苦的一个月的时候，他就像获悉了他以后将经受这样的一个月那样的苦恼。对于那些令人愉快的事件他同样中立。当他被告知说他以后将享受某个非常愉快的时期的时候，他对获悉这一点感到欣然。他极为展望这个时期。当人们提醒他说他曾经享有过这样的一个时期的时候，他同等地感到欣然。我将称这个想象的人为无时间者（Timeless）。

这个人与我们非常不同。但是他的描述是前后一致的。我们能够因而拒斥上述的提法。我们或许缺乏未来偏向这一点是可以想象到的。即便我们在时间上不会完全中性，但是我们本来也可以会像无时间者一样。

67.我们为什么不应当偏向未来

我们的未来偏向对我们来说是糟糕的。如果我们像无时间者的话，对我们而言就会更好一些了。我们在某些方面会有所失。于是，当糟糕的事情处于过去的时候，我们不应当感到如释重负。但是我们也应当有所得。当好事处于过去的时候我们不应当悲伤。

所得在权重上会超过所失。一个理由会是这样的。当我们回顾的时候，我们能够足以承担起选择性。当这有助于帮助我们避免重蹈覆辙的时候，我们应该牢记我们的生活中的某些糟糕的事情。但是我们会允许自己忘记已经发生过的绝大多数糟糕的事情，而通过反复重温来保留我们所有的美好记忆。在我们展望的时候如果这么具有选择性，对我们而言则会是糟糕的。除非我们考虑到所有可能发生的糟糕事情，否则我们就失去避免这些事情的机会。既然当我们展望的时候不应该具有选择性，但是在回顾的时候我们却能够承受得起这么做，那么后者总体上说来会更加令人愉快一些。[23]

会有其他一些更大的收益。其一会是在我们对待变老和死亡的态度之中。让我们首先来考虑这样一个论证，伊壁鸠鲁借以主张我们未来的非存在不能够是某种痛惜之事。我们并不痛惜我们过去的非存

在。既然如此，我们为什么应当痛惜我们未来的非存在？如果我们安之若素地对待一个，那么我们为什么不把这个态度延伸到另一个？

一些人主张说，这个论证是失败的，因为我们可能活得更长久，但是我们本不能够更早出生。这并非一个好的诘难。当毕达哥拉斯学派得悉2的平方根不是一个有理数的时候，他们对此感到痛惜。甚至当一些真理是假的这一点在逻辑上是不可能的时候，我们也会对这些真理感到痛惜。

伊壁鸠鲁的论证之所以失败另有原因：我们具有未来偏向。因为我们具有这一偏向，所以只是认识到我们一度受苦这一点现在可能并不烦扰我们。但我们安之若素并不表明我们过去所受之苦就不是糟糕的。就我们过去的非存在而言同样会如此。因而伊壁鸠鲁的论证只有对那些既缺乏未来偏向、又不痛惜他们过去的非存在的人们而言才具有说服力。既然没有这样的人，那么该论证就不对什么人具有说服力。

尽管该论证失败，但是它可能提供某种安慰。如果我们害怕死亡的话，该论证表明我们所惧怕的对象不是我们的非存在，而只是我们未来的非存在。想到我们过去的非存在我们能够处之泰然，并不表明那不是某种令人痛惜的事情。但是，既然我们事实上并不心怀恐惧地看待我们过去的非存在，那么当想到我们不可避免的死亡的时候，我们可能能够用这个事实来减少我们的恐惧或者沮丧。如果我们时常想到而且泰然自若地对待我们身后的黑暗，那么这种安之若素可能转移到我们对待我们身前的黑暗的看法上。

现在假设我们缺乏未来偏向。我们就像无时间者一样。那么我们应当会在我们对待衰老和死亡的态度方面大大获益。随着生活的推移，我们可以展望的东西应当会越来越少，而可以回顾的东西会越来越多。如果我们设想另一个差异，这个后果将更加明显。假设我们的生活不是随着出生和童年开始，而是像亚当那样开始。假设，尽管我们都是成年人，而且具有成年人的见识和能力，但是我们只是刚刚开

始存在。我们缺乏未来偏向。我们会被我们昨天并不存在这个想法大大地烦扰吗?

这取决于非存在错在何处。一些人认为它本身就是糟糕的。但是更加令人信服的观点认为，它惟一的过错是它造成我们的损失。假设我们接受这个观点。那么我们可能把它当作一个根据，据以对我们的生活是一个两端受到非存在约束的有限生活而感到痛惜。但是，如果我们刚刚开始存在，我们不会认为某种糟糕的东西就在我们身后。我们据以感到痛惜的根据只会是我们已经错过了许多本来会很好的东西。假设，即使我生为1700年前后少数享受特权的人物之一，我也仍然会像我现在实际上的这个样子。那么我会极为痛惜我事实上是出生在1942年。我会极为宁愿经历过先前两个半世纪的生活，在我的朋友中拥有像休谟、拜伦、契诃夫、尼采和西季威克这样的人物。

在我的那个想象出来的情形中，我们并不偏向未来，而且我们刚刚开始存在。尽管我们会痛惜我们没有更早一些存在这个事实，但是我们不会为昨天我们还并不存在这个想法大大地烦扰。我们不会带着绝大多数现实中的人们对待明天突然死亡这一前景会有的那种恐惧或悲伤来对待这个事实。我们之所以不会具有这样的恐惧或者悲伤，是因为，尽管我们不会有什么好东西可以回顾，但是我们会有可以展望的整个生活。

现在假设我们的生活几乎已经过完。我们将在明天死去。如果我们并不偏向未来的话，那么我们的反应就应当反映出我刚刚表述的那个反应。我们不应当为我们很快将不再存在这个想法大大地烦扰，因为，尽管我们现在没有什么可以展望的，但是我们拥有可以回顾的整个生活。

可能会遭到这样的诘难：“你现在能够回顾。但是，一旦你死了，你就不能够回顾了。而你明天将死去。所以你应当会为此而大大地烦扰。”我们本来能够回答说：“为什么？的确，在我们不再存在之后我们将永远不能够享受对我们的生活的回顾。我们现在没有什么可以回

顾的，甚至没有回顾之快乐。但是同等为真的是，在我们开始存在之前，我们不能够享受展望我们生活的快乐。在我们刚刚开始存在之后，我们根本没有什么可以回顾的，甚至没有展望之快乐。但是那时我们没有什么受到巨大烦扰的理由，因为那时我们能够展望我们的整个生活。既然我们现在能够回顾我们的整个生活，为什么那个平行的事实——我们没有什么可以展望的——带给我们受到巨大烦扰的理由？”

这种推理忽略了那些本质上指向未来的情绪。它不会适用于其展望之快乐来自制定计划或体验到诸选择的那些人们。但是当应用于那些坐享到来的生活快乐、更为被动的类型的人们的时候，该推理看起来是对的。而且，鉴于这对于我们而言部分是对的，那么这个推理表明，如果我们缺乏未来偏向的话，我们本会更加幸福。我们因衰老和死亡的临近所感受到的沮丧本会少得多。如果我们本来像无时间者的话，那么处于我们生命的终点就会更像是处于生命的起点了。在我们生活中的任何一点，我们本可以享受到对我们的整个生活的回顾或展望。

我业已主张，如果我们缺乏未来偏向的话，这对我们而言本来会更好一些。这与这样一个令人信服的主张相匹配，那就是，如果我们缺乏近期偏向的话，这对我们而言本来会更好。只是批判后一个偏向的话，在此并没有什么根据。这两个对待时间的态度总体上说来对于我们都是糟糕的。

既然我相信这个态度对于我们是糟糕的，那么我认为我们不应该偏向未来。这个信念并不是以有关这个偏向的合理性的待证假定作为推论的根据。按照任何一个令人信服的道德观点，如果我们全都更加幸福的话，会更好一些。正是在这个意义上，如果我们能够的话，我们不应该偏向未来。在给予我们这个偏向方面，进化否定了我们对待死亡的最佳态度。

68.时间的流逝

回到我的主要问题。这些对待时间的态度是非理性的吗？我们绝大多数人认为未来偏向并不是非理性的。我们倾向于认为，缺乏这个偏向会是非理性的。如是，我们可能并不完全被那个想象出的事例中所给出的推理所折服，在那个事例中我们是时间上中立的，而且明天将死去。我们能够描述某个这样的人，他并不怎么在意明天将死这个前景，因为他现在能够回顾他的整个生活。但是这个态度，尽管是可以描述的，可能看起来有些怪诞，或者包含一种荒谬的错误。

举一个更为简单的事例将有所帮助，其中不包含非存在和我们对待一个整体生活的那些态度。这可能是早先的一个例子的变体，涉及我们所想象出来的时间上中立的人。考虑一下

> 无时间者如何迎接好消息。无时间者在医院接受一个痛苦的手术，之后将伴随着一种诱发性失忆。他醒过来，对先前的一天没有什么特定的记忆。他问护士他将在什么时候不得不忍受这个痛苦的手术以及要忍受多长时间。如前所述，那位护士知道有关两个病人的事实，但是不敢肯定他是哪一个。不过，无论是两种情况中的哪一个，他的手术都需要非常长的时间，要足足持续10个小时。该护士知道下述一种情形为真。他要么昨天遭受了10个小时的痛苦，要么今天晚些时候将遭受10个小时的痛苦。
>
> 无时间者陷入忧伤。他本来希望手术更短一些。
>
> 当该护士返回的时候，她欢呼说："好消息！你是那位昨天遭受过痛苦的病人"。
>
> 无时间者仍然愁眉不展。"这为什么是好消息？"，他问道。"我的煎熬一样痛苦，一样长。而且它同样是我生活的一部分。我的煎熬是在过去这一点为什么就应当给我现在造成什么差异？"

对于这个事例而言所诱发的失忆可能构成对它的一个诘难。我因而添加

> 事例二。无时间者做了这个手术，而且没有失忆。我们在他遭受煎熬的前一天和后一天去拜望了他。在后一天，无时间者还是那样愁眉不展。"为什么我应当感到宽慰？"他问。"我所受的煎熬是在过去怎么就会更好一些？"

无时间者是在犯一个错误吗？他应该感到宽慰吗？我们绝大多数人会给出肯定的回答。但是，倘若不以待证假定为论据，就难以解释为什么。我们或许说：“如果煎熬处于他的未来的话，他就会仍然不得不去经受它。既然它处于他的过去，那么它就过去了，了结了。”这不是对于为什么无时间者是非理性的这一点的一个进一步的解释。他“仍然”不得不去经受该痛楚，我们只是重申了它处于未来。

我们在此或许可以诉诸所谓的时间的流逝，或时间流变的客观性。我们或许说：“如果他的痛苦处于未来，那么它将越来越近，直到他实际上遭受该痛苦。但是，如果他的痛苦处于过去，那么它将只是越离越远。”这样的说法看起来表达了一个深刻的真理。但是这个真理却令人好奇地难以捉摸。“它将越来越近”意思指什么？难道这不是仅仅意味着，在未来的那些时刻，这个未来的痛苦将越来越临近将成为那时的当下时刻的东西吗？但是，在过去的那些时刻，一个过去的痛苦更靠近彼时的当下时刻。这种非对称性处于何处？

在回应中使用一个特定的隐喻——穿越时间的运动——是自然的。我们可能说我们是在穿越时间进入未来，或者未来的事件是在穿越时间进入当前，抑或当前或曰“现在”的范围是在进入未来。“现在”沿着历史事件的序列下移，“就像聚光灯沿着一排合唱队的女孩在下移。”

把“现在”（now）与“在此”（here）加以比较可能有所帮助。对于那些否认时间的流逝或时间流变的客观性的人们而言，“在此”和“现在”是严格地具有类比性的。它们两者都相对于一个特定思想者的思想或者话语。“在此”指的是这个思想者在某个时间所处的地点，而“现在”指的是某个特定的思想——牵涉到“现在”概念的一个思想——被思考到的时间。这两个词本可以被“此”（this）所取代，就像在宣告者的行话“此时此地”中那样。[24]

那些相信时间的流逝的人们会拒斥这个类比。他们会承认，在一个不包含什么思想者的宇宙当中，概念“在此”会缺乏应用。但是他们

主张，即便在这样的一个宇宙当中，某些事情现在正在发生仍然会是真的，然后其他事情现在正在发生会是真的，然后复有其他事情现在正在发生会是真的，等等。甚至在一个没有生命的宇宙当中，“现在”的范围仍然会穿越时间从过去进入未来。

这个穿越时间的运动的隐喻可能是无法辩护的。我们穿越时间的运动有多快？我们可能不会满足于那个惟一可能的回答——“以每秒一秒的速率”。我们可能主张，如果要么是我们、要么是“现在”能够穿越时间运动，那么这个运动更快一些或更慢一些必定是有意义的，除非这没有什么意义。

该隐喻的那些批评者可能证明是正当的。但是这可能并不表明就没有时间的流逝或时间流变的客观性这样一回事。这大概是一个“绝对真理”（a categorical truth），处于这样深的一个层次，以至于我们不应当期望它能够以隐喻的方式或者以其他什么方式得到解释。[25]

我不想竭力断定在这个辩论中真理位于何处。因而我对这两种可供的选择都将予以考虑。首先假设，就像一些哲学家和物理学家所认为的那样，时间的流逝是一种幻觉。如果是这样，时间中性就不能够是非理性的。在捍卫自利论方面，S论者必定谴责近期偏向。如果时间中性不能够是非理性的，那么S论者可能回到他早先的观点，即这样的中性是合理地所需要的。他然后必定主张，就像在某个无可避免的痛苦被推迟了的时候感到如释重负是非理性的一样，当某个无可避免的痛苦处于过去的时候就感到如释重负同样是非理性的。我们将会发现这是难以置信的。

其次假设，我们相信时间的流逝或时间流变的客观性是对的。那么S论者可能维持他后一个观点，而且诉诸时间的流逝。他必定仍然谴责近期偏向。他可能主张说：“尽管你有绝好的理由较少关心他人的痛苦，但是你不能够合理地较少关心你的处于更加远的未来中的痛苦。与当前时刻的纯粹距离在合理性上并不重要。”S论者现在可能以一种不同的方式支持这个主张。他可能放弃诉诸时间上的中性，即那

个有关纯粹的时间选择在合理性上并不重要的主张。他可能代之以区别对待不同种类的时间关系。

我们此处应当谨记的是，我们绝大多数人具有第三种对待时间的态度：当前偏向。如果纯粹的时间上的选择在合理性上并不重要，那么更加关心当前的痛苦就不能够是合理的。我现在处于极度痛苦之中并不能够构成现在更加关心这个极度痛苦的根据。这可能看起来荒唐。当应用于当前偏向的时候，时间中性的要求可能看起来最不令人信服。当我在遭受极度痛苦的时候更加在意我的这个极度痛苦如何能够是非理性的？这样的一个主张看起来从根本上削弱了整个的关心的结构。只是因为当我们现在痛苦的时候所感受到的，痛苦才重要。只是因为未来的痛苦将在未来成为当前的痛苦，我们才关心未来的痛苦。如果未来的那些痛苦的表现就像"爱丽丝的明天停顿"[26]一样，而且永远处于未来的话，它们就会一点也不重要了。[27]

S论者现在可能主张说："就我们对待时间的三种态度而言，一个是非理性的，另外两个则是理性所要求的。我们必须（must）更关心当前的痛苦，而且我们不能够（cannot）合理地关心过去的痛苦，但是我们切勿较少关心那些处于更远的未来而不是处于较近的未来的那些痛苦。"这个新观点缺乏一般性之吸引力。而时间选择上的纯粹差异——对"何时？"问题的纯粹回答——在合理性上并不重要这个主张中却有颇具吸引力的简单性。这个新观点尽管不那么简单，但是仍然可能证明是正当的。S论者或许主张，一经反思，它是直觉地令人信服的。他或许主张说："当我们比较当前、过去和未来距离的时候，非常明显的是，前两者完全不同于第三者。前两者显而易见在合理性上很重要，为我们在关心方面的差异提供了正当证明。但是第三者显然是微不足道的。"

这些直觉不具有普遍性。就那些当无可避免的糟糕事件被推迟之时感到如释重负的人们的情况而言，许多人并不认为这种如释重负之感是非理性的。要么考虑一下近期偏向的另一个后果：随着某个美好

事件临近当前，我们所感到的那种高涨的兴奋——就像在剧院中的观众席上的照明灯暗下来的那一刻所感到的那样。许多人会把这样的兴奋看成并非是非理性的。

S论者或许说：“那些具有这些直觉的人们还没有充分考虑该问题。那些已经考虑过该问题的人们，诸如那些哲学家们，一般都同意这样一点，那就是，更为关心较近的未来是非理性的。”

正如我已经说过的，哲学家们的一致同意可能并不证明他们的观点正当。自利论长期以来处于主导地位。既然S已然被传授了两千多年的时间，那么我们必定期盼在我们的直觉中发现其某种回音。 S不能够仅凭诉诸可能已经由其教诲所孕育出的那些直觉而得到正当性的证明。

如果时间的流逝不是一种幻觉，那么S论者则无需只是诉诸我们的直觉。他能够主张，时间的流逝证明未来偏向是正当的。如果被要求解释为什么，他可能发现这相当困难。例如，没有什么迹象表明过去是不真实的。如果过去并不真实的话，那么就会非常容易地明白为什么过去的痛苦不重要了。为什么因为时间的流逝过去的痛苦就并不重要，则并不如此显而易见。

S论者或许主张说：“假设我们允许‘现在’的范围移动到未来这样一个隐喻。这解释为什么三种时间态度中的一个是非理性的，而另外两个则是理性所要求的。痛苦之所以重要，只是因为当它们处于当前或者处于‘现在’的范围的时候它们像什么。此乃我们现在处于痛苦之中的时候为什么必须更为关心我们的痛苦的原因。‘现在’移动到未来。此乃过去的痛苦为什么并不重要的原因。痛苦一旦过去，它们将只是远离‘现在’的范围。事情因未来的近之特性而有所不同。时间的流逝并不证明更为关心近期的未来是正当的，因为无论未来的痛苦多么遥远，它们终将进入‘现在’的范围之内。”

这些论证是否是些好的论证并不清楚。特别是最后一个可能是以待证假定作为论据。但是S论者可能代之以主张，诉诸时间的流逝，

我们无需论证。他可能主张同样无需进一步的解释。这一点可能是另一个基本的真理，那就是，既然时间流逝，那么过去所受之苦完全不会是重要的——不会是合理关心的对象。获悉他的煎熬是在过去，无时间者并未感到如释重负。这可能不涉及能够得到解释的那种错误。该错误可能是如此重大，以至于超出了论证所及的范围。

69.一种非对称

通过放弃诉诸时间中性，而代之以诉诸时间流逝，自利论者大概已经加强了他的地位。但是我们应当考虑最后一类事例。我称之为我们所爱的那些人过去或未来所受之苦。

> 事例一。我是来自某个国家的流亡者，我的寡母还留在那个国家。尽管我对她极为关心，但是我很少得到有关消息。一段时间以来我知道她得了致命的疾病，将不久于人世。我现在获悉一些新情况。我母亲的疾病已经变得非常痛苦，而这种痛苦药物无法解除。在她死前接下来的为数不多的月份中，她将面临极度的煎熬。她将很快死去这一点我已经了解。但是得悉她必须忍受痛苦我还是感到深深的忧伤。
>
> 一天之后我被告知，我所获得的消息部分不准确。有关事实是对的，但是时间方面不对。我的母亲的确受了数个月的苦，不过她现在已经死了。

我现在应该感到如释重负吗？我原以为我母亲所受的煎熬是在未来。但它却是在过去。根据S论者的新观点，过去的痛苦一点也不重要。获悉我母亲所受之苦这一点并未赋予我现在感到忧伤的理由。现在就宛如她是毫无痛苦地死去一样。如果我仍然悲伤的话，我就像那位无时间者一样了。我就是在犯一个如此重大的错误，以至于它超出了论证所及的范围。

最后这个例子可能对S论者有所撼动。他可能发现难以相信我的反应是非理性的。他可能说：“你母亲是否经受那些受苦的月份对你而言怎么可能是重要的？即便她经受了，她所受之苦也是在过去。这根本不是什么坏消息。”但是，在应用于我对别的什么人的关心的时候，这些说法就显得较少有说服力了。

该S论者可能代之以试图修正他的观点。他可能说："我本不应当主张过去的痛苦简直是不重要。时间的流逝所蕴含的是，它们的重要程度较少一些。"但是这个修正是无法得到辩护的。一旦一个痛苦过去，它就完全过去了。处于过去不是一个重要程度问题。提出这样的主张不能令人信服，那就是，既然时间流逝，那么对于过去的痛苦具有某种关心、但又少于对未来痛苦的关心是合理的。有关我过去的那些煎熬应当作何主张？就这些情形而言，我对我过去所受之苦完全漠然置之。这是非理性的吗？尽管比我对于未来所受之苦所感到的哀伤要小一些，但是我应该有些哀伤吗？诉诸时间的流逝不能够令人信服地支持这个主张。而且难以相信，在这些情形中我的漠然是非理性的。

我所列举的事例揭示出我们在关心自己的过去和他人的过去方面的一种令人惊讶的非对称性。如果有人提醒我说我自己曾经一度不得不忍受数个月的痛苦的话，我一点也不会感到哀伤。但是，如果我获悉我的母亲在死前不得不忍受这样的一种煎熬的话，我会极为哀伤。

这种非对称性在下面的事例中有所减小：

> 事例二。此事例除了这样的一点之外——尽管我的母亲遭受了数个月的痛苦，但是她仍然活着，而且现在并不痛苦——就像事例一一样。

获悉我母亲过去所受之苦，此处我会较少哀伤一些。这种差异能够得到解释。如果我的母亲像我，她现在对她过去的煎熬漠然置之。（我们能够假设她对这个煎熬的记忆本身并不痛苦。）如果我们在关心自己和他人过去所受之苦方面有一种非对称性，那么倘若这种非对称性在他人现在已经死亡的那些情形中最为明显的话，则不会令人惊讶。如果我的母亲仍然健在，我当前的态度将自然而然地受到我所能设想到的她的当前态度的影响。既然我能够设想她现在对她过去所受之苦漠然置之，这就可能减少我对这个苦难的关心。但是，如果我母亲现在死了，那么她现在并不对她过去所受之苦漠然置之。既然我对

她过去所受之苦的关心不能够受到她的当前态度的影响，那么这就是我的关心以其最纯洁的形式示人之时。

她所受之苦是否持续到她的死亡这一点造成什么差异吗？

考虑一下

事例三。我得悉我母亲遭受了数月的苦，但是在她死前有一个月没有痛苦。在她的生活之内有这么一个时期，其中她所受之苦是在过去，而且如此一来对她而言不再重要。

如果这就是我所获悉的，这会给我的关心造成什么差异吗？我相信它最多会造成一点差异。获悉我母亲受了那么多月的苦我会深深地哀伤，即便我也知道她有一个月所受之苦处于过去。令我感到哀伤的不只是获悉我母亲痛苦之死。如果只是这点使我哀伤的话，而且获悉她死前不得不忍受数月的大量痛苦我并不哀伤的话，我的反应就会是如此地特别，以至于它恐怕会被忽略。但是我对那些我所爱的、现在已逝世的人们的那些过去的关心，不仅仅是对他们并未痛苦地死去的关心。获悉他们在生活中的任何时刻遭受了我原先并不知情的数月的痛苦，我都会感到哀伤。我相信绝大多数人在这个方面同我一样。

最后考虑一下

事例四。此事例除了这样的一点——既然我那时就知道了，那么我就并非获悉我母亲所受之苦——之外，其他则与事例三一样。

纵使我已经总有这一认识，但是一想到我母亲的生活中有数个月充满苦难仍然会令我黯然神伤。我再次相信，类似的主张适用于大多数其他人。就我们对待我们自己过去所受之苦而言仍然存在显著的非对称性，对于我们自己过去的苦难我们绝大多数人都近乎漠然置之。

可能有诘难说：“如果我们作出区分，那么这种非对称性就消失了。你问当苦难在过去的时候它是否重要。这串联起一些不同的问题。你是否应该感到同情是一个问题，你是否应该有所关心则是另一

个问题。苦难是否在过去并不给同情造成差异，而只是对关心造成差异。我们只对他人感到同情。此乃你为什么对你过去的苦难漠然置之的原因。你不能够同情你自己。当你得悉你母亲的苦难的时候，你确实也应该感到同情。但是担心这个过去的苦难则会是非理性的，就像担心你自己过去的苦难会是非理性的一样。”[28]

我认为，这些主张并未消除这个非对称性。在事例一一开始，我就被告知，我母亲在死之前将遭受数月的痛苦。一天之后这个信息得到纠正：她死前确实遭受了数月之苦。按照刚刚陈述的那些主张，在早一天我应当大为担心，其时我相信我母亲的苦难将是在未来。当我获悉它是在过去的时候，我应当不再担心，尽管我仍然应当感到同情。当我不再有任何担心的时候，这总该大大地减轻我的哀伤吧，而且还改变其性质。但是我肯定的是，如果这个想象出来的事例出现的话，我的态度不会以这两种方式发生改变。我的哀伤或许较少一点，但是这个差异不会很大。我的哀伤也不会改变其性质。

某个事件是否在过去会影响到、也应当影响到我的那些与可能的行动相连的情绪。但是在这些事例中，当我认为我母亲的苦难是在未来的时候，我做什么都不会有用。我甚至不能够给她捎信。我因而不能够具有那种实际的关心——寻找我能够用以帮助我所关心的那个人的一些方式。在这些事例中，我的关心只能够是消极的。它只能够是忧愁和哀伤，没有寻求可能的补救方法的冲动。因为我的哀伤会呈现为这种形式，所以当我获悉我母亲的苦难是处于过去的时候其性质不会改变。

我承认，当我获悉这个事实的时候，我的哀伤或许较少一点。正如，如果我母亲活着的话，我的关心或许受到我母亲态度的影响一样，它或许也受到我对待我自己过去所受之苦的态度的影响。这个影响可能部分消除该非对称性。在我对自己所受之苦的关心之中，无论这个苦难是在过去还是未来，它都会造成差异。如果有关我的态度的这个事实影响到我对那些我所爱的人们的过去的苦难的关心的话，不

会令人感到吃惊。既然我对他们的过去苦难的关心不可避免地受到我对我自己的过去苦难的关心的影响，那么我对他人所受之苦的关心决不能够呈现为完全纯粹的或者没有扭曲的形式。而且，正如我已经主张的，当我获悉我母亲的苦难是在过去的时候，我的关怀之心不会大大减少。

按照上面的诘难，我对我自己过去的苦难没有什么关怀之心，是因为一个人不能够同情他自己。这个主张一点也没有消除该非对称性。它只是一种重述。它承认了在我们对待我们自己生活中的过去苦难与我们所爱的那些人的生活中的那些苦难的态度之间有这个差异。

这一非对称性使捍卫自利论更为困难。一个S论者不能够令人信服地主张这个非对称性是合理地所要求的。特别是，他不能够在此令人信服地诉诸时间的流逝。如果时间的流逝证明我对我自己过去的苦难完全漠然置之为正当，甚或使这种漠然置之成为一种合理的要求，那么该S论者对于我对我所爱的那些人的关心必须有相同的主张。在有关我母亲之死的那个想象出来的事例中，她的苦难是在过去同样为真。

该S论者有关我们对于过去之苦难的态度应当主张什么？他可能主张说：“此处不存在一个独一无二地合乎理性的态度。如果你对你自己过去的苦难完全漠然置之，这并非是非理性的。但是对你自己过去苦难的了解导致你大为哀伤的话，这也不会是非理性的。同样，如果对你母亲过去苦难的了解使你大为哀伤的话，这不会是非理性的。但是如果你对她的苦难完全漠然置之的话，这也不会是非理性的。”

如果该S论者承认这一系列对过去的不同态度不是非理性的，他如何能够捍卫其有关在我们对未来的关心方面我们应当时间中性的主张？他仍然必须作此主张。但是，如果在过去所受之苦的情形中，同样地关心，或较少地关心，抑或一点也不关心并不会是非理性的，为什么在未来所受之苦的情形中只有一个态度是合理的？尽管没有彻底

的不一致，但是难以相信在有关一系列对待时间的态度方面如此宽容的一个观点却在有关另一系列态度方面如此地严苛。

70.结论

我得出的结论是，一个自利论者有望能够予以辩护的观点只有两个：

（1） 如果时间的流逝是一种幻觉，那么时间上中立的就不能够是非理性的。该S论者或许重新提出我们必须时间上中立的主张。那么他必须主张，当苦难已经被推迟的时候，以及在它在过去的时候感到如释重负两者都是非理性的。如果他批评近期偏向，那么他也必须批评未来偏向。如果时间的流逝是一种幻觉，那么他必须同意（a）缺乏未来偏向不会是非理性的。他不能够也主张（b）具有这个偏向并不是非理性的，以及（c）具有近期偏向是非理性的。没有什么论证他可以用来支持这3个主张。如果他不谴责未来偏向，那么他就不能够凭借有关近期偏向对我们而言是糟糕的这个主张，来谴责近期偏向。未来偏向对于我们而言也是糟糕的。而且一个态度的合理性并不取决于它对我们而言是否糟糕。在这两个对待时间的态度之间存在一种差异：尽管我们能够直接按照近期偏向而行动，但是我们不能直接按照未来偏向而行动。但是这并不能够支持只有第一个偏向是非理性的主张。该S论者不能够主张，因为我们不能够按照未来偏向而行动所以它不是非理性的。如果他诉诸时间上中立，那么他因而必须主张，当我们的苦难在过去的时候感到如释重负是非理性的。我们将发现这难以置信。

（2） 如果时间流逝不是一种幻觉，那么该S论者可能辩护一个与之不同的观点。他或许主张，因为时间流逝，过去的苦难不会是重要的。那么他能够主张，对于无时间者而言当他获悉他的苦难完结了的时候并不感到如释重负则是非理性的。当我们思考我们自己的过去的时候，我们将发现这个观点令人信服。但是该S论者也必须主张，当

我获悉我母亲在死之前受了苦的时候感到哀伤是非理性的。我们将发现这难以置信。

该S论者自己可能发现最后这个主张难以置信。如果他放弃这个主张，那么他就必须放弃诉诸时间的流逝。尽管这个诉诸或许支持那个笼统的有关过去的苦难简直不重要的主张，但是它并不能够支持那个有关我们被合理地要求对于过去的苦难要具有某种却较少的一些关心的主张。它也不能够表明在我们对待自己过去的痛苦和他人过去的痛苦的态度方面的区别是合理的。

即使时间的流逝不是一种幻觉，该S论者也可能回归他的第一个观点：要求时间上中立。那么他能够借助那个有关时间选择上的纯粹差异在合理性上并不重要的主张来谴责近期偏向。他能够主张，尽管何人感觉到某个痛苦是合理地重要的，但是某个痛苦何时被感受到则不能够是重要的。

如果他回归这个观点，该S论者必须谴责当前偏向。正是此处时间上中立看起来最不令人信服。当我处于极度痛苦的时候我们更加在意我的极度痛苦如何能够是非理性的？该S论者或许说：“在某种意义上，这不是非理性的。极度痛苦之所以是糟糕的，只是因为那时你多么在意它。但是在另一个意义上，你不应当偏向当前。如果让这样的一个偏向影响到你的决定会是非理性的。尽管你那时更在意该极度痛苦，但是你不应当，因为这一点，以预见到的后来更大的极度痛苦为代价而终结你当前的极度痛苦。在第一阶层面上，当你在感受到它的时候你更加在意你的极度痛苦。但是你不应当更加关心它是处于当前而不是处于未来。在第二阶的层面上，在你作出影响到你所受之苦的长度和时机的决定之处，你能够而且应当是时间上中立的。”

如果他是在要求时间上中立，那么该S论者也必须谴责未来偏向。他可能说：“我们应当期待这个偏向由进化所产生。这解释为什么这个偏向只是诉诸或者更多地诉诸我们自己的生活。当我们考虑他人

的生活的时候，我们能够上升到我们的进化的遗传性之上，而且能够看到时间上中立的令人信服之处。”

当某个信念或者态度具有一种进化的解释的时候，这本身就具有中立的蕴含。它本身不能够表明这个信念或者态度要么得到了正当证明，要么没有。但是假设我们具有其他一些根据来置疑某个态度。其辩护者们或许说：“我们全都具有这个态度这一事实就是认为它得到正当证明的一个根据。如果它并未得到正当证明的话，它为什么得到这么广泛的支持？”在回答这个问题的时候，一种进化的解释可能给它所解释的投下一层怀疑的阴影。它从根本上削弱了那个对立的解释——我们具有该信念或者态度是因为它得到正当证明。该S论者因而可能主张，我们未来偏向在我们自己的生活中是进化的一个纯粹产物，而不是合理地得到正当证明的。而且这个主张看起来得到在我们对我们自己过去的生活和他人过去的生活的关心方面的那个非对称性的支持。[29]

该S论者会不得不把这个主张应用于“我过去的或者未来的那些手术”。在这些事例中，我想要我昨天的确遭受了数个小时的痛苦而不是我今天晚些时候将遭受一个小时的痛苦为真。该S论者必定再次主张，这个优先选择是非理性的，而且一般说来当我们的苦难在过去的时候感到如释重负则是非理性的。即便有鉴于他有关进化的新观点，我们将发现这仍然难以置信。

我已经表述了该自利论者能够最令人信服地予以辩护的两个观点。这些观点中的每一个观点都至少包括一个难以置信的主张。这是自利论的一个弱点。在这些观点之间存在一种选择又是一个更大的弱点。较少关心更为遥远的未来，这一点可能是非理性的。但是在为什么这么认为的理由尚属悬而未决的情况下，对于此点我们则还不能够肯定。

---

[1]Nagal（1），第74页。

[2]参见Williams（3），第215页：

……像对他人的深深依恋这样一些事情……不能够同时体现那个不偏不倚的观点，而且……它们还冒着违反它的风险。如果它们存在，它们就冒着那个风险；然而除非这样的一些事情存在，否则一个人的生活中就将没有足够的本质或者确信以迫使他忠诚于生活本身。任何一个事物如果要有意义，包括执着于那个不偏不倚的系统，生活就得具有实质；但是如果它具有本质，那么它就不能够给予那个不偏不倚的系统以至高无上的重要性，而据其成立的那个系统在这个限度上将是不可靠的。

[3]参见Hare（4），第159—167页，以及第175—184页。

[4]Nagal（4），第十四章。

[5]Bentham第四章第二段。

[6]Lewis（1），第483页。

[7]Hume（1）第三卷第二章第七节。

[8]这个图表是我经由R·诺齐克从G·安斯利那里获取的。

[9]参见Strotz。

[10]Plato（1）。

[11]Pigou，《福利经济学》，伦敦，麦克米伦出版社，1932年，第23页。

[12]Hume（1）第三卷第二章第七节。

[13]》Hume（1）第二卷第三章第三节。

[14]Rawls，第420页。

[15]该事例的这个版本是由G·哈曼向我提出的。

[16]Rawls，第293页。

[17]Hume（2），第177页。

[18]我在重复R·诺齐克的一些评述。

[19]济慈（John Keats，1795—1821），英国最伟大的诗人之一。其作品音律优美、古意盎然。代表作有《夜莺颂》、《秋颂》和《希腊古瓮》等。——译者

[20]凡·高（Vincent van Gogh，1853—1890），荷兰后印象主义派著名画家。早期作品用色较暗；后期作品则用色明亮。代表作有《向日葵》、《农民》、《邮递员罗兰》等。——译者

[21]艾丁提出了这个诘难的一个更加微妙的版本。我并未完全答复艾丁的诘难。

[22]Anscombe（1），第68页。

[23]我认为这些获益本身会压倒那位S论者在第65节开始之处所描述的那些损失。

[24]这些评述过于简单化。即便我们否定时间的流逝，我们也能够承认"现在"能被用于对一个无生命的宇宙的描述。如果两者同时发生，那么事件X"现在"就是相对于事件Y。

[25]参见Pears（1），第249页：

而且，因为时间是一个范畴，而且不显眼地出现在我们所有的经验之中，所以用很多不同的方法来定义过去和未来是可能的，而且这些定义之中没有哪一个是非常令人信服的。就给予时间性词汇以逻辑关系的那些句子而言，它们展示出一个令人好奇的特征。它们是……可以说是，弱同义反复。而且它们之所以是弱同义反复，不仅是因为我们如此习惯于正确使用时间

性词语以至于我们无需什么强提示，而且是因为它们的结构是特有的。
就绝大多数同义反复而言，就像大理石石柱是通过一个大理石石鼓垒放在另一个之上构成的一样，它们是直接堆砌术语而构成的。但是赋予时间性词语以逻辑关系的那些同义反复则是把它们的术语像那些拱顶石那样砌在一道。没有什么单独的术语关联是不可或缺的，或者是能够独立的。但是它们一道形成我们在上面描画知识壁画的那个穹顶。
对于时间流逝的讨论参见Gale（2）（特别是威廉斯和古润堡姆的文章），Gale（1），以及Smart（2）。

[26]爱丽丝的明天停顿（Alice's Jam Tomorrow）：作者在此采用了英国作家刘易斯·卡罗尔的《爱丽丝漫游奇境记》中"明天停顿，昨天停顿，但不是今天停顿"一语的一个短语。此短语被人们用来比喻永远处在未来的、无法兑现的承诺。——译者

[27]这些评述太粗略。即便是在当前，一些经验是中性的，甚或是不悦人的，但是能够非常快活地予以牢记。Proust（1），第97页写道：
……在一个晚得多得多的日子，当我以相反的顺序来渐次回味我在那么深地爱上阿伯丁妮之前所经历的那些时刻的时候，当我的那颗伤痕累累的心能够没有痛苦地从死去的阿伯丁妮身上脱开的时候，我得以没有痛苦地详细回忆起了那一天，阿伯丁妮与弗朗索瓦去购物而不是呆在特罗卡德罗的那一天；我带着快乐忆起它，就像属于那时我才知道的一个道德季节；我的回忆详细而精确，没有给它添加任何现在的痛苦，而是相反，就像我们追忆夏季的某些日子，当那些日子持续着的时候我们发觉它们太热，只有在它们过去之后我们才从之萃取毫无杂质的精金和永远不灭的蔚蓝。
Proust（1），第107页：
爱的印象与生活的那些其他印象是不成比例的，但并非是当它消失在它们当中的时候我们才能够说明它。不是从它的脚下，不是在喧闹的街道和麇集的房屋当中，而是当我们远离的时候，从附近的山坡上，从整个镇子已经消失了的一个距离，或者显得就像地上的一团乱物的一个距离，我们才能够，在超然的孤独和薄暮中，鉴赏一个大教堂的顶点，独一无二，坚忍而纯粹。
即便一个人从未在当前从他的任何一个经验中发现任何快乐，但是过一种具有巨大的后顾的和前瞻的幸福的生活会是可能的。

[28]这个诘难是由J．Broome、 R．Swinburne和J．Thomson向我提出来的。

[29]参见Singer（2）。

第九章

我们为什么应当拒斥自利论

一个自利论者必定谴责近期偏向。其中一个诘难是说这个偏向是非理性的。正如我们刚刚看到的，这个诘难并非一帆风顺。

71.诉诸后悔莫及

该自利论者或许诉诸一个与众不同的诘难。他可能对重近者说：

> 你现在并不痛惜你的近期偏向。但是你将会感到痛惜的。当你付出代价的时候——当你遭受到你以使之更糟为代价而推迟了的那个痛苦的时候——你将但愿你未曾更多地关心你的较近的未来。你将后悔你有这个偏向。去做你明知将会感到痛惜的事情是非理性的。

如前所述，这个诘难是不准确的。当重近者付出代价的时候，他可能后悔过去有过近期偏向。但是这并不表明他必定后悔现在有这个偏向。相似的主张适用于那些利己的人们。当一个利己的人付出他人的利己行动所施加在他身上的代价的时候，他痛惜这些他人是利己的这一事实。他痛惜他们的自身偏向。但是这并不导致他痛惜他自身中的这个偏向。在人际层面上的有关S的真理适用于在时际层面上的P。正如一个利己的人所痛惜的不是他自身中的他的偏向而是他人中的他的那种偏向一样，重近者所痛惜的不是他自己现在自身中的他的偏向，而只是他在其他时刻自身中的他的偏向。当我假定了重近者并不痛惜他的偏向的时候，假定他并不痛惜他当前的偏向就足够了。这是他总是照此行动的偏向。上述的诘难并未表明重近者必定痛惜这个偏向。

我们接下来能够注意到，重近者不会总是痛惜他的过去的偏向。在较近的过去，现在作为其当前的与其近期的未来的东西都是近的，所以那时他表现出有利于它们的偏向。既然他现在正得益于而且将有

一段时间得益于这个过去的偏向，他现在将为他有过这个偏向而高兴。

该自利论者可能说："当你付出代价的时候，你将痛惜你过去的偏向。既然如此，那么你现在甚至就应该痛惜你当前的偏向。你未来将痛惜你当前的这个偏向；而且现在你关心你的未来。关心一个人的未来包括但愿避免这个人将会痛惜的东西。既然你将痛惜你当前的这个偏向，你现在就应该但愿你未曾有过这个偏向。"

重近者会回答说："在更远的未来我的确将痛惜我当前的这个偏向。而且这的确赋予我但愿我现在并未具有这个偏向的诸根据。但是在较近的未来，我将对我有过我当前的这个偏向而高兴，因为那时它将有益于我。这赋予我为现在具有这个偏向而高兴的诸根据。"

该S论者或许说："你在近期未来中的高兴将被你在更远未来中的痛惜所超过。这将是真的，因为你以预见到了的、它们变得更糟为代价推迟了这些痛苦。既然你的当前偏向后来将造成你的痛惜多于高兴，你但愿现在不具有这个偏向的那些根据，在两组相互竞胜的根据中，是较强的那一组。"

重近者会回答说："不偏不倚地看来，我未来的痛惜的确将超过我未来的喜悦。但是我并不不偏不倚地考虑我的未来。我更关心近期是什么。既然我未来的喜悦更近，那么它超过了我未来的痛惜。"

该S论者可能回应说："并非不偏不倚是非理性的。"重近者会回答说："你的这个回应要么是无后果的，要么是自杀性的。如果它击败我，那么它也击败你。那些利己的人们并非不偏不倚。正如我有近期偏向一样，他们有着对自身的偏向。"

72.重近者的失败何以不是自利论的胜利

在辩护其近期偏向方面，重近者把这个偏向与那个对自身的偏向加以比较。S论者或许说：

我较不关心他人身上发生什么。你较不关心后来那些时候在你身上将发生什么。所有这些后来的时候将在某个时刻成为对你而言的一个现在。但是其他人将永远不是对我而言的一个我。你的类比因而是失败的。[1]

这个诘难有某种力量。假设，因为这个或者其他一些理由，我们拒斥重近者所主张的。假设，尽管有第八章中所出现的那些困难，我们得出近期偏向是比那个对自身的偏向较少合理性的结论。正如我已经解释的，这并不表明我们应当采纳S。近期偏向是自利论者们所喜欢打的靶标。在我尝试辩护这个偏向方面，我是在这个关键点上挑战S。如果这个尝试取得成功，S就被完全击败。但如果这个尝试失败，并不必然是S获胜。

当前目标论的最佳版本是那个批判论的版本（CP）。正如我写到的，CP能够主张我们是被合理地要求去关心我们自己的利益的，而且是以一种时间上中立的方式。按照CP的这个

版本，重近者是非理性的，因为近期偏向是非理性的。如果我们相信重近者是非理性的，那么这决不是什么接受S而不是CP的这个版本的理由。

73.诉诸不一致性

假设，正如CP的这个版本所要求的，我以一种时间上中立的方式关心我自己的自我利益。但这不是我的主要的关心。因为我还有其他一些有时更为强烈的愿望，我有时以我明知有悖于我自己的自我利益的一些方式行动。S论者可能重新诉诸以后的痛惜。他说："因为你的行动有悖于你的自己利益，你未来的痛惜将超过你未来的喜悦。与重近者不同的是，你对你的整个未来给予同等关心。你因而应当承认，你的行动是非理性的。去做你明知总体上说来将会感到痛惜的事情是非理性的。"

这个诘难假定，每当我的行动有悖于我自己的自我利益的时候，我以后都将痛惜这个行动。这个假定没有得到正当证明。我之所以以这样一些方式行动，是因为，尽管我关心我自己的自我利益，但是我却更加关心别的事情。既然我并不是最关心我自己的自我利益，那么我们就不应当假定我后来将痛惜这些行动。

很有可能存在这样一些情形，其中我的确痛惜某个过去的行动。但这并不必然得出，我的行动是非理性的。假设我之所以有悖于我自己的自我利益而行动，是因为我接受过一个我现在予以拒斥的价值判断。我现在可能痛惜这个行动，因为我已经改变了我的主意。但是我的行动可能仍然一直是合理的，因为我那时是在按照其时我所接受的一个价值判断而行动。

当我的愿望在不改变主意的情况下发生变化的时候，相似的主张也适用。假设，在过去，我之所以有悖于我自己的利益而行动，是因为我想要帮助那些落难的人们。我借了一大笔钱来帮助这些人，而且明白，要还清这笔借贷我将不得不到我所痛恨的一个行当中干上许多年。我现在想要帮助另一些落难的人们。我并不认为这些人有更大的主张得到帮助的权力。但是，因为我现在更加生动地意识到他们的不幸，以至于这些人成为我现在最想给予帮助的人们。因为我早先的行动，我现在不能帮助这些人。鉴于我尚未付清贷款，我不能够再借钱来帮助这些人们。我因而痛惜我先前的那个行动。该行动有悖于我的利益；而且我现在对此感到痛惜。但是，有鉴于我那时最想要的，它则并非是非理性的。诉诸以后的痛惜可能表明重近者是非理性的。但是它绝不是什么对我现在正在讨论的那个CP版本的诘难。

根据CP的这个版本，我应当具有对我自己的未来的一种时间上中立的关心，但这无需是我的那个占主导地位的关心。而且，就像所有版本一样，CP的这个版本赋予我们当前的目标以特殊的权重。它因而可能受到内格尔针对较粗糙的工具论所陈述的那个诘难的挑战。内格尔写道，根据这个理论：

我会有理由现在去做那些将肯定造成我未来的合理企图落空的事情；我会有理由去做我明知以后将有理由努力予以取消的事情，因而我将不得不特别小心地在我未来自我的道路上设置一些陷阱和难以逾越的障碍。带有这样的一些后果的一个体系不仅未能要求时间过程中行动方面最为基本的一致性，而且通过在合理性的官能中确立个体图谋针对他未来的自我的根据，事实上加剧了那些冲突的可能性。[2]

内格尔所描述的"不一致性"并非理论上的不一致性。在有关某个人做什么是合理的问题上，工具论并不是在不同的时候持有一些不一致的主张。也无需某个认为这个理论或批判性的当前目标论的人置疑别的时候他自己的合理性。这是因为，按照P的所有版本来说，理由不仅是相对于行动者的，而且也是相对于行动时间的。

的确，根据P，我现在去做那我以后予以取消的话将会合理的事情能够是合理的。 P能够遭到这样的指控，那就是，即便在一个单一的生活之内它也是时际间自败的。甚至就CP的那个要求我们以时间上中立的方式关心我们自己的自我利益的版本而言，这也能够为真。假设我就有这样的一个关心，但是我还有其他更加强烈的愿望，而且这些愿望在不同的时刻有所不同。在任何一个时刻，我要么能够（1）做那将最好地实现我的当前愿望的事情，要么能够（2）做那将最好地实现或者最好地使我能够实现我毕生的所有愿望的事情。根据P，我总是应当做（1）而不是（2）。正如我们在第34节中所看到的，有这样的一点能够是真的，那就是，如果我总是遵循P，做（1）而不是（2），那么我甚至在实现我每一次的愿望方面都将历时地较少成功。

一个自利论者可能在此主张，甚至以P的标准来衡量S也会击败P。在其愿望实现论版本中，S告诉我们总是要做（2）而不是（1）。如果我总是以这种方式行动，那么，我每一次在实现我的愿望方面将历时地更加成功。

我解释过这如何能够得到回答。对S也有一个相似的诘难。根据S，对我来说去做那对你而言予以取消的话反而合理的事情能够是合

理的。正如P能够是时际自败的一样，S能够是人际或者集体性自败的。一个自利的人们的共同体如果全都遵循某个道德版本，而不是自利论的话，他们甚至以自利的标准衡量也会做得更好。但是就S的情形而言集体性自败并不是毁灭性自败。当S集体性自败的时候，它仍然是个体性成功的。既然S是一种有关个体合理性的理论，那么以其自身标准衡量它仍然是行得通的。

还需要我写这一段吗？就P的情形而言，时际自败并非毁灭性自败。当P是时际自败的时候，它在每一次还是成功的。仍然为真的一点是，如果我在每一次遵循P，那么我在每一次就是在做将最好地实现我的那些当前愿望的事。甚至在这些情况下，从行动时的行动者的观点看来P也是成功的。既然这就是P所诉诸的观点，那么它以其自身的标准衡量仍然是行得通的。

这两个诘难不能够驳倒P或者S。但是它们都具有某种力量。与P不同的是，S不能够是直接地时际自败的。这赋予S较之于P的一种特定的理论吸引力。而且这可能促使我们放弃P而接受S。但S能够是直接地集体性自败的。就一个行动者中性的道德论而言则不是如此。这赋予这样的一种道德论以相似的理论吸引力。而且这可能促使我们采取相似的进一步的步骤，就是从S到这样的一种道德论。

鉴于两个诘难之间的类比性，对P的那个诘难并不支持S。这两个诘难都支持中立主义。如果这些诘难成功，那我们就应当拒斥S。如果这些诘难失败，我们就没有理由拒斥P。[3]

一个S论者或许否认该类比性具有这些蕴含。他可能主张，尽管一个可以接受的理论不能够是直接地时际自败的，但它能够是直接地人际自败的。如何能够有这种差异？ S论者必定主张，不同人们之间的关系在那些相干的方面不同于某个时刻的一个单一的人与在另一个时刻的他自己之间的关系。

这些关系在绝大多数方面是不同的。我与你的关系不同于现在的我与明天的我或者50年之后的我之间的关系。但是这些关心在诸相关

的方面仍然可能是相似的。就像一个行动必定是某个特定的行动者的行动一样，它也必定是在某个特定的时刻做出的行动。况且，许多有关合理性的主张只有在应用于一个处于某个特定时刻的人的时候才为真。当它们被弄得既横跨不同的人们之间的关系，又横跨一个人的此时与彼时之间的关系的时候，它们就不再为真。于是乎可能有主张说："一套不一致的信念能够合理地被不同的人们所相信，但是它们不能够合理地被一个单一的人所相信。"但实际并非如此。一套不一致的信念可能合理地被一个单一的人所相信，如果这个单一的人是在不同的时候相信它们的话。只有当他在单一的时刻相信它们的时候，他才是非理性的。就非传递性的优先选择而言——宁取X而舍Y，宁取Y而舍Z，以及宁取Z而舍X——也是如此。正如被人们时常指出的那样，3个不同的人每一个能够具有这些优先选择中的一个，而并非是非理性的。但是，如果他在不同的时间具有这些优先选择，一个单一的人也能够如此。正如这些主张所提示的，当我们是在考虑理论上的合理性和实践上的合理性这两者的时候，一个现在的人与其他时候的他自己之间的关系是与不同的人们之间的关系具有相关的相似性的。

74.结论

在第六章到第八章中我提出了几个针对自利论的论证。这些论证证明两个结论之一是正当的。该组论证可能只是表明S不能够击败P。按照这个结论，这些理论之间的辩论以平局而告终。当S与P冲突的时候，无论遵循两者中的哪一个都会是合理的。但是，正如我在附录B中所解释的，这个结论在实践上对S来说会是一个失败。

另一个结论是说我们应该拒斥S。无论是在实践上还是在理论上，这对S来说都会是完全的失败。（如果我们拒斥S，就像我在附录C中所解释的那样，这可能影响到我们在那些有关自我利益的不同理论之间的选择。）

我认为我的那些论证证明这个更大胆的结论是正当的。我以一个战略性的隐喻开始。自利论有两个对手：道德论和当前目标论。在某些方面它是处于这两个对手之间。因而它时常容易受到致命的攻击：它会遭到来自两个方向的攻击。自利论主宰我们的理智传统由来已久。但是这种主宰地位主要源于它的两个对手未能联手发起攻击。当受到道德论者们的攻击的时候，它则借力于当前目标论，反之亦然。

我从这两个方向上都对自利论提出了挑战。这点确保了S只会按其自身的是非曲直而受到评判。我避免了一个人的作为确实只影响到他自己这样一个欺骗性的情形。 S告诉这个人去做会对他自己而言最好的无论什么事情。既然他是他的行动所影响到的惟一的人，他就是在做那对每一个受到影响的人而言最好的事情。他是在做不偏不倚地考虑下来具有最好后果的事情。在这样的一种情形中，S与不偏不倚的仁爱相契合。当这两者相冲突的时候，亦即当对行动者而言更好的事会以更大的差幅对另一些人而言更糟的时候，S能够更好地得到评判。正如这些情形所表明的，S固守一个具有偏向性的关心模式。 S并不仅仅是谨慎，而且是自我主义。它坚持的是，一个合理的行动者必须赋予他自己的自我利益以至高无上的权重，他人付出无论什么代价都不足惜。

然后我从另一个方向对S提出了挑战。我考虑了其中S与当前目标论相冲突的一些情形。在这些情形中，尽管行动者明知那些事实，而且思虑清楚，但是他并不想要给予他自己的自我利益以至高无上的权重。 S主张说，无论他人的代价是什么，即使在冷静下来的时候他既不具有也不想要具有这个偏向，一个合理的行动者也必定带有对自身的偏向。

这个主张可信吗？这个偏向是独一无二地合理的还是至高无上地合理的？每一个别的愿望或者关心都较少合理一些吗？这是个主要的问题。我的第一论证的回答是否定的。我主张说，较之那个对自身的偏向，还有几个别的愿望绝不较少合理一些。例子之一就是行动要符

合他人的利益这样一个愿望。即便当一个人明知其行动有悖于他自己的自我利益的时候，去实现这个愿望也能够是合理的。其他一些例子包括某些种类的成功的愿望。一个创作者可能想要他的那些创作尽可能地好。一个科学家或者哲学家，可能想要作出某些根本发现，或者推进理智进步。我主张，这些和其他别的一些愿望并不比那个对自身的偏向较少合理性。如果其中的一个是某一个人的最强烈的愿望，在所有的事情都考虑到的情况下，对他来说造成它得以实现会是合理的，即使这个人明知他的行动有悖于他自己的自我利益。

S论者的第一回应与这些主张相矛盾。这个回应主张说，那个对自身的偏向是至高无上地合理的。既然我不能够证明我拒斥这个主张是对的，那么我的第一论证就不是决定性的。但是我相信它是成功的。我相信S论者对于这个论证没有什么好的回应。那个对自身的偏向不是至高无上地合理的。至少有一个愿望——给他人带来益处的愿望——绝不较少合理性。既然至少有一个愿望是这样的，那么我们就应当拒斥S，并接受CP的某个版本。

S论者的第二回应诉诸任何理由的力量都是在时间中延展开来的这一主张。按照这个主张，既然我将有一些理由去努力实现我未来的愿望，那么我现在就有这些理由。那些行动理由不能够相对于一个特定的时刻。对这个主张的一个论证可能还表明那些行动理由不能够是相对于行动者的。该论证可能表明，任何一个理由的力量都在不同的人们的生活中展开。如果内格尔的论证成立，这就是它所表明的。这个结论既会击败S又会击败P。要避免这个结论，S论者必须主张理由能够是相对于行动者的。我主张说，如果理由能够是相对的，那么它们能够是相对于行动时刻的行动者的。正如我在第59节至第61节中所表明的，我的确具有过或者我将具有某些行动理由这一点能够为真，尽管我现在并不具有这些理由。这从根本上削弱了S论者的第二回应。而我诉诸完全的相对性则进一步为拒斥S提供了根据。

S论者还必须主张，较少关心一个人更远的未来是非理性的。第八章表明，S论者要主张这一点则必须接受两个观点之一，其中的每一个都有一些难以置信的蕴含。这是对S的另一个诘难。出于论证的目的，我假定了这个诘难能够得到答复。我假定了较少关心我们更远的未来是非理性的。这并不表明我们应当接受S。我们本可以接受当前目标论的那个批判性版本。而CP能够主张，我们被合理地要求去以时间上中立的方式关心我们自己的自我利益。该主张并没有把这两个理论区分开来。

S要求我们接受一个更加大胆的主张。我们具有这个对自身的、时间上中立的偏向并不够。无论他人的代价是什么，即使我们既不具有也不想要具有这个偏向，我们也必定总是受到这个偏向的支配。这个主张把我们带回到那个主要问题。根据我的第一论证，这个主张要求这样的假定，那就是，这个偏向是至高无上地合理的。它要求这样的假定，即更多地关心别的任何事情，诸如道德或他人的利益，都是非理性的。我们应当拒斥这种假定。如果S论者没有什么别的回应，我们就应当拒斥S。

S论者还有另外两个论证：诉诸以后的痛惜和诉诸不一致性。尽管这些论证具有某种直觉魅力，但是它们并未提供对我的第一论证的回应。我得出我们应当拒斥S这一结论。正如我所预见到的，自利论不能够在它的对手——当前目标论和道德论——两者的联合攻击下死里逃生。

当前目标论的最好版本是那个批判性版本。下面谨记的一点是，如果我们接受CP，我们本可以主张我们最强烈的愿望是避免错误地行动这一点是被合理地要求的。至于我们是否应当把这个主张添加给CP，我则未下结论。既然如此，那么有关合理性的幸存着的理论就不是两个。道德论者应当接受CP。他们没有什么根据来拒斥CP。因为CP能够把道德论者们认为那些道德理由所应该有的所有权重赋予这些理由。

最后谨记，有关合理性的所有的可能的理论都是CP的一些版本。因为此点为真，所以无论什么是我们所相信的，我们都应当接受CP。CP的这个特征可能看起来是个弱点，使之成为一个虚空的理论。但这个特征是个强点。当不同的理论重新表述为CP的一些版本的时候，我们就能够更加清晰地看清它们所假定的是什么。而且，尽管我对CP关于道德理由应当主张什么未下结论，但是我并未对另外两个问题悬而未决。考虑一下休谟的追随者们，他们否定愿望要么能够是固有地非理性的，要么是合理地所要求的。如果我们给CP添加这个主张，那么它则与IP，即纯粹的工具论相契合。我业已主张，我们应当拒斥CP的这个版本。某些关心模式是非理性的，并不提供什么行动理由。而且我的主要主张是，我们应当拒斥CP的那个与S相契合的版本。我们应当拒斥这一假定，那就是，较之那个对自身的偏向其他每一个愿望都是较少合理性的。假设我们的愿望和价值判断既不是单独地又不是集合地非理性的。而且假设我们明白，将会最好地实现这些愿望的东西将有悖于我们自己长远的利益。如果是这样，那么，遵循S则是非理性的。当我们知道这将使我们在知道事实且思虑清楚的情况下最想要的和最珍视的东西落空的时候，我们还去做符合我们自己的自我利益的事情则是非理性的。

自利论两千多年以来一直受到绝大多数人的信奉。正因为如此，迫不及待地主张我们应当拒斥S可能看起来有些荒唐。区区四章的篇幅如何能够推翻有记载的历史的裁决？这么多的人如何能够一直是错的？答案有二。

（1） 这些人中的绝大多数假定，因为我们会有来生或者会转世，所以道德论和自我利益总是相契合的。正是因为这些人有这样的虚假信念，所以他们忽视了对S的一个诘难。

（2） 正如我们应当拒斥某个理论的时候经常有的情况一样，那些相信这个理论的人们也不是完全错的。 S的某些部分是令人信服。更多关心自己并不是非理性的。而且在我们关心我们自己的自我利益

方面，我们大概应当是时间上中立的。这些主张的令人信服之处有助于解释为什么这么多的人们已经相信了自利论。但这些主张也是CP这个我们全都应当接受的更加广泛的理论的组成部分。考虑一下这个（过于堂皇的）类比。牛顿定律部分是正确的。但是我们现在接受一个与之不同的理论。

---

[1]诺齐克向我提出的这个诘难当应用于那个对未来之偏向的时候，其力度要小一些。

[2]Nagel（1），第40页。

[3]尽管它们具有某种力量，但这两个诘难不应当被过誉。 S并非指令所有的人发动一场针对所有的人的战争，理由至少有三。即便那些纯粹自私的人们的利益在非常大的程度上也是相重叠的。即便当这些利益发生冲突的时候，S通过产生我所称的政治解决之道，本身也可能告诉这些人们努力化解这个冲突。（当该事例涉及充分少的人的时候，S就将告诉这些人们这一点。）而且绝大多数自利的人并不完全是自私的，而是对他人也有某种关心。
所有这三点在事例P中都有各自的类比物。如果某人知道相关事实而且思虑清楚，他在不同的时刻最会要的东西在非常大的程度上会是重叠的。甚至当这一点不为真的时候，P本身告诉我们各人要力图化解“时际的格斗”——努力与其他时刻的他自己达成政治的解决方案。如果有两种方法某人现在可以用来努力实现他当前的愿望，而且其中之一更有可能干扰到他后来实现其未来愿望所做的努力，P则告诉这个人以另一种方式行动。即便在有人对他自己的未来没有什么关心的那些极端情形中，在我们诉诸P的那个非批判性的版本——工具论——的情况下，这也会是真的。正如我从G·哈曼处所学到的，即便在这个情形中，如果这个人选择了将不会干扰到他的未来努力的那个行动方式，在达到他现在想要的东西方面也将更加成功。他之所以将更加成功，是因为后来他将没有理由干扰他的当前努力的后果。其三，绝大多数行动者事实上关心未来那些时刻的他们自己。而且批判性的当前目标论可能以一种时间上中立的形式要求这样的关心。


第三编 个人同一性
上海译文出版社

第十章

我们认为自己是什么

我进入电子传输器。以前我去过火星，然而却只是借助那种老的方式——花费数周时间乘坐太空飞船旅行的方式——前往的。而这部机器将以光速把我送达火星。我只要按一下那绿色的按钮。像其他人一样，我感到紧张。能行吗？我提醒自己人们所告知我的预期的情形。当我按下按钮时，我将失去知觉，然后似乎在片刻之后会苏醒过来。而事实上我将失去知觉大约1个小时。在记录我的所有细胞的精确状态的同时，地球上的这个扫描仪将毁坏我的大脑和躯体。然后它将通过无线电信号传输该信息。信息以光速运行，将用3分钟的时间到达火星上的复制器。然后这部机器将用新的物质创造出与我的大脑和躯体精确相似的一个大脑和躯体。我将在该躯体中苏醒过来。

尽管我认为这就是将会发生的事情，但还是有些犹豫不决。好在接着我记起了今晨早餐时，我对此流露出紧张情绪的时候，妻子粲然一笑的情景。当时她提醒我说，尽管她经常被电子输送，她还不是好好的。我于是按下按钮。正如所预期的，我失去知觉而且好像立刻重获知觉，但却身处一个不同的舱室中。审视着自己的新身体，我发现一点也没变。甚至早晨刮胡子时在上嘴唇上留下的那个小伤口也仍然在那里。

几年过去了，期间我时常被电子传输。现在我又一次步入那个舱室中，准备再一次到火星上去。但是这一次当我按下绿色按钮时，我没有失去知觉。先是出现了嗡嗡的声音，然后是一片寂静。我离开舱室，对服务员说："机器没有正常工作。我的操作有什么不对吗？"

"机器工作正常，"他一边回答我，一边递给我一张打印出的卡片。卡上写着："新扫描仪记录下了你的图像，没有毁坏你的大脑和躯体。我们希望你会欢迎这个技术进步所带给你的机会。"

服务员告诉我说，我是首批使用新扫描仪的人之一。并且补充说，如果我逗留1小时的话，我就能够用内部通话设备看到并与火星上的我自己对话。

"先等一下，"我回答说，"如果我在这里，我就不能也在火星上。"

有人轻轻地咳嗽一声。这是一个身穿白大褂的人，他要与我私下谈谈。我们来到他的办公室。他让我坐下，停顿了片刻后说："恐怕我们的新扫描仪遇到了一些问题。它就像你与火星上的你自己交谈时所看到的那样精确地记录了你的图像。但是看来它有损它所扫描的对象的心脏系统。就迄今的结果来判断，尽管你将在火星上相当健康地生活，但是在这里，在地球上，你不日将难逃心力衰竭的厄运。"

后来服务人员把我叫到内部通话设备旁边。在内部通话设备荧屏上，我看到了自己，就像每天早晨在镜子中看到自己一样。但是有两个不同之处。我在荧屏上不是左右反相的。而且，就在我站在这里默默无言时，我能够看到并且听到我自己在火星上的工作室内开始对我讲话。

从这个想象的故事我们能够学到什么？一些人认为我们从中学不到什么。如果碰到维特根斯坦的话，[1]他会持这种观点。而奎因写道：“科幻小说的方法在哲学中有其用武之地，但是……不知道该方法的限度是否引起了人们的适当注意。在前所未有的环境下去为个人的相同性寻求‘逻辑上所要求的’东西，就是在提示语词具有某种超出我们过去的需要所赋予它们的逻辑力量。”[2]

如果在考虑到这样的想象事例时我们毫无反应的话，这一批评本来会得到正当的证明。但是这些想象事例在我们绝大多数人中激起了强烈的信念。而且这些信念之为信念，不是关于我们的语词，而是关于我们自己。我们通过对这些事例的考虑，我们发现，我们所相信的东西涉及我们自己的连续存在，或者发现使现在的我们与明年的我们是同一些人的东西。我们发现了我们关于时间中的个人同一性的本质的信念。在考虑想象的事例时我们的信念固然得以最清晰地表露，可是这些信念也涵盖一些实际事例，以及我们自己的生命。在本书的第三部分，我将先论证这些信念中有一些信念是虚妄的，进而提示这如何重要，以及为何重要。

75.简单电子传输和支线事例

在我的故事之初，扫描仪毁掉了我的大脑和躯体。我的图像用电子束发射到了火星，那上边有另一部机器制作出了我的一个有机复制品。我的复制品认为他是我，而且他看起来记得我直到按下绿色按钮时的那一刻的生活。在其他每一方面，不论在身体上还是在心理上，我们都确切地相似。如果他返回地球的话，每一个人都会认为他就是我。

就像刚刚描述的那样，简单电子传输是科幻小说中的一个常见特色。对某些读过这种小说的人们而言，据认为这是最为快捷的旅行方式。他们相信我的复制品会是我。另外一些科幻小说读者，以及这类小说中的某些人物则持一种不同的观点。他们认为，当我按下绿色按

钮的时候，我死了。我的复制品是某个别人，只是已经被制作得与我确切地相似而已。

第二种观点看起来得到了我的故事结局的支持。新扫描仪并没毁坏我的大脑和躯体。除了采集信息之外，它仅仅损坏我的心脏。在我处于舱室中的时候，随着绿色按钮被按下，好像什么都没有发生。我走出来，并且获悉在数天后我会死亡。我随后通过双向电视系统，与火星上的我的复制品进行了通话。让我们来续写这个故事。由于我的复制品知道我将要死去，他就用我最近用来安慰一个垂危朋友的那些同样的方法来努力安慰我。得悉这些想法在接收端多么没有安慰后果真是件伤心的事情。然后我的复制品向我再三保证他将从我生命停止的地方延续我的生命。他爱我的妻子，并会一道照顾我的子女。而且他将完成我正在写的这本著作。除了拥有我所有的草稿之外，他还具有我所有的意向。我必须承认，他能够像我一样好地来完成我的著作。所有这些事实对我起到一点安慰作用。在我知道自己将有一个复制品的情况下慢慢死去，并不太像简单地死去那么糟。即便如此，我还是将很快失去知觉，永远地失去知觉。

在简单电子传输中，我在被复制之前遭到摧毁。这使人们更容易相信这是一种旅行方式——我的复制品是我。在我的故事的结尾，我的生命和我的复制品的生命出现了重叠。不妨称之为支线事例。在该事例中，我不能希望在主线上旅行，带着等在前头的40年寿命在火星上苏醒过来。我将呆在数天后终结的支线上，即呆在这里的地球上。既然我可以同我的复制品通话，那么看来很清楚：他不是我。尽管他与我确切地相似，但他是一个人，我则是另一个人。当我捏自己的时候，他没有什么感觉。当我的心脏发病时，他还是没有什么感觉。而且当我死去时，他将再活40年。

如果我们相信我的复制品不是我，那么假定我在该支线上的前景几乎像平常的死亡一样糟，就是很自然的事了。我将否定该假定。就像我后面将要论证的，被摧毁和被复制这一点差不多与平常的存在一

样好。我能极好地为该主张辩护，而且在讨论过有关个人同一性的以往辩论之后，能极好地为包容了该主张的一个更广的观点辩护。

76.质的同一性与量的同一性

相同或同一性有两类。我与我的复制品是质上的同一，或者说确切地相似。但是我们可能不是量上的同一，或者说可能不是同一个人。同样，两个白色的撞球不是量上的同一，而可能是质上的同一。如果我把这些球中的一个涂成红色，它将停止在质上与过去的它自己同一。但是我后来所见到的红球和我涂上红色的那个白球在量上是同一的。它们是同一个球。

就某个人而言，我们或许说，"在发生事故之后，他不再是同一个人了。"这是有关这两类同一性的一种看法。我们宣称他——那同一个人，现在不是同一个人。这不是一个矛盾。我们只是指这个人的性格发生了变化。这个量上同一的人现如今在质上是不同的。

当我们关心我们的未来时，我们所关心的是我们量的同一性。我可能认为婚后我将不再是同一个人。但是这并不使婚姻死亡。不论我怎么变化，我将仍然活着，如果有某个将是我的人活着的话。

尽管我们主要关心的是我们的量的同一性，但是心理的变化是重要的。的确，照一种观点，某些种类的质变摧毁量的同一性。如果某些事情发生在我身上，真相可能不是我变成了一个非常不同的人。真相或许是我不再存在——结果存在的是别的某个人。

77.个人同一性的物理准则

一直存在着有关个人的本质和有关时间中个人的同一性的本质的大辩论。它将有助于区分这些问题：

（1）一个人的本质是什么？

（2）什么使一个处于不同时间点上的人成为同一个人？什么是必然包含在每一个人在时间中的连续存在的东西？

对问题（2）的回答可以采取这种形式："X今天与在过去某个时刻的Y是同一个人，当且仅当……"这样的一个回答陈述了时间中个人同一性的充分和必要条件。

在对（2）的回答中我们还将部分回答（1）。我们的连续存在的必要特征有赖于我们的本质。而对（1）的最简单回答是，要做一个人，必须是一个有自我意识的存在，必须对它的同一性及其在时间中的连续存在有所意识。

我们还可以问

（3）在每一个人时间中的连续存在实际上包含了什么？

既然我们的连续存在中有并非必然的一些特征，对（2）的回答只是对（3）的部分回答。例如，具有同一的心脏和同一的性格对我们的连续存在而言并不是必要的，但是它们通常是这种存在所包含的部分。

许多学者使用"时间中的同一性准则"这一模棱两可的短语。一些人用来指"我们用以辨别当前某对象是否与过去某对象相同一的方法"。但是我将用来指这同一性所必然涉及的，或者所寓于的。

就绝大多数物理对象而言，照我所称的标准观点，时间中同一性的准则是该物体的时空物理连续性。这是我们大家都理解的，即使我们不理解我将要作的描述。就最简单的物理连续性事例而言，比如埃及金字塔，一个显然静态的物体持续存在。就另一个简单事例而言，比如月亮，一个以一种有规则的方式运动的物体。很多物体的运动方式不那么规则，但是它们仍然有物理上连续的时空轨迹可循。假设我涂成红色的那个台球撞球与去年我拿下取胜分的那个白色的台球撞球是同一个球。照标准观点，只有当该球循着这样的一个连续轨迹时这才为真。它在下述条件下一定为真：（1）有一条贯穿时空的线路，始自我打出获胜一击之前那个白球所停在的地方，止于现在那个红球

所在的地方，（2）在该线路的每一点上都曾经有一个台球撞球，而且（3）在该线路每一点上一个球的存在部分地是由先前紧接的一点上一个球的存在引起的。[3]

某些类型的事物甚至在它们的物理连续性包含着巨大变化的情况下继续存在。一种叫“Camberwell Beauty”的蝴蝶最初是虫卵，之后是毛虫，再后是茧蛹，最后才是真正意义上的蝴蝶。这些是一个单一有机体物理上连续存在的四个阶段。其他一些类型的事物不能在这样巨大的变化之后仍然幸存。假设一个画家画了一幅自画像，然后对之重新加工，变为他父亲的一幅画像。尽管这两幅肖像之间的相似之处要远远大于一条毛虫和一只蝴蝶之间的相似，但是它们不是某个单一物体的连续存在中的阶段。自画像作为一幅画被艺术家毁了。在对同一性的一般讨论中，我们会需要解释，对不同类的事物而言为什么物理连续性的需要有这样的一些不同。但是我们这里对此可以按下不表。

一个物理对象的连续存在中能够有间隙吗？假设：即便我儿时得到的一块金表一度被拆成零件在修表匠的架子上躺了一个月之久，我仍然拥有那同一块金表。照某一种观点，就该表所循的时空轨迹而言，不是在每一点上都存在一块表，所以我的表并不具有一个完全物理连续性的历史。但是在我的表被大卸八块、（作为表）并不存在的那一个月期间，它的所有的零件是具有完全连续性历史的。照另一种观点，即使当我的表被肢解时，我的表也还存在。

另一个复杂情况还是涉及到一个复合物件与它的各种各样的构成要件之间的关系。尽管就所有这些事物而言情况并非都为真，但是其中某些事物的连续存在不需要包含它们的构成成分的连续存在却是真的。假设一条漂浮在港湾中的木船不时进行修理，而且五十年后船身上不再留有当初造船所用的哪怕一点点木头。但是它仍然是那同一条船，因为作为船，它在五十年的风雨中展示出所有的物理连续性。尽管事实上它现在是由完全不同的一些木头构成的，但是此点仍然成立。这些木料在质上或许与原初的木料同一，但是它们不是同一些木

料。就人的躯体而言部分地也与此相类似。除了某些脑细胞例外之外，我们身体的细胞在我们的一生中都经历了数次更新。

我现已描述了照标准观点看来，使一个物理对象在经过许多天甚至许多年之后仍是同一事物的物理连续性。这使我能够选择一有关个人同一性的敌对观点予以陈述。照这观点，正是我有同一个大脑和躯体这一点使我在时间中是同一个人。我的时间中的同一性的准则，或者说这同一性所包含的，是时间中我的头脑和躯体的物理连续性。我将继续存在，当且仅当这特定的大脑和躯体既继续存在又是一个活生生的人的大脑和躯体。

此为该种观点的最简单版本。还有一个更好的版本。这就是

> 物理准则：（1）必要的不是整个躯体的连续存在，而是足以使大脑之为一个活生生的人的大脑的连续存在。今天的X与过去某个时刻的Y是同一个人，当且仅当（2）Y的大脑的足量的连续存在，而且现在是X的大脑，而且（3）该物理连续性没有呈现为一种“分支”形式。（4）时间中个人的同一性就寓于（2）和（3）的事实的成立之中。

在某些实际事例中（1）清清楚楚地为真。一些人继续存在，即便他们丧失了他们躯体的许多部分或者功能。（3）将留待以后予以解释。

那些相信物理准则的人们会拒斥电子传输。他们会以为这不是一种旅行方式，而是一种赴死方式。他们还会因难以想象而拒斥转世重生。他们认为某人不能够在死后还有生命，除非这种生命处在同一物理连续躯体的复活体中。这就是为什么某些基督教徒坚持土葬的理由。他们相信，如果他们像希腊和特洛伊英雄们一样被焚化，骨灰四散，那么，即使上帝也不可能再让他们死而复生。这样的话上帝只能创造一个复制品，即某个确切地与他们相似的人。另一些基督教徒相信，如果上帝用他们最后活着时构成他们躯体的物质一点一点地把他们重新组装起来，上帝能够复活他们。这就像重组我的金表。[4]

78.心理准则

有些人相信一种类似物理连续性的心理连续性。这包含一个纯粹精神实体或曰事物——灵魂或曰心灵本质——的连续性存在。我将返回这一观点。不过现在我首先解释另一类的心理连续性。这与物理连续性的相似之处更少，因为它并不寓于某种实体的连续性存在之中。但是这另一类的心理连续性只涉及到我们所熟悉的事实。

被讨论得最多的向来是记忆的连续性。这是因为正是记忆使我们绝大多数人意识到时间中我们自己的连续存在。患有失忆症的人们是些例外。绝大多数的失忆症患者只丧失两套记忆。他们失去对过去特定经验的记忆——（我们为用语简短起见）或曰经验记忆。他们还失去关于事实的记忆，即那些有关他们自己过去生活的一些记忆。但是他们记得其他事实，而且他们记得如何做不同的事情，诸如如何说话、如何游泳等。

洛克提出，经验记忆提供了个人同一性的准则。[5]尽管就其本身而言，它尚不是一个看似合理的观点，但是我相信它可以是这样一种观点的成分。因而我将回应洛克的一些批评家们。

洛克主张，除非某个人现在记得他犯过罪，否则不可能犯过罪。我们能够理解对那些不记得所犯罪行的人们加以惩罚的踌躇。但是，就其被当作有关什么是包含在一个人的连续性存在中的东西之观点来看，洛克的观点明显是虚妄的。如果洛克的观点为真的话，对某人而言，忘记他曾经做过的任何事情或者他曾经有的任何经验，就是不可能的了。但这是可能的。我现在记不得今晨穿什么衬衫。

有数种展开经验记忆标准的方式，以便涵盖这样一些事例。我将诉诸经验记忆重叠链这一概念。比方说，在今天的X和20年前的Y之间存在着直接的记忆联系，如果X现在能够记得Y在20年前具有的经验。照洛克的观点，只有这使现在的X与20年前的Y是同一个人。但是，即使没有这样的直接记忆联系，在现在的X和20年前的Y之间也可能有记忆的连续性。如果在现在的X与那时的Y之间一直有一个重叠的直接记

忆链的话，就会这样。就绝大多数成年人而言，会有这样的一个记忆链。在过去的20年间的每一天，他们这些人中的绝大多数都记得在前一天的一些经验。照洛克观点的修正版本，某个现在的人X与某个过去的人Y是同一个人，如果在他们之间有记忆的连续性。

这个修正应对了对洛克观点的一个诘难。我们还应该修正这个观点，以便它能诉诸其他事实。除了直接记忆，还有数种别的直接心理联系。其一就是一个意向和实现该意向的后来行动之间的联系。其他这样的直接联系还有当连续持有一个信念、愿望或者其他任何心理特征时所具有的那些联系。

我现在能够定义两个一般关系：

心理上的联系性支持着特殊的直接心理联系。
心理上的连续性支持着重叠的强联系性链条。

就这两个一般关系而言，联系性在理论和实践上都更为重要。联系性可以有不同的程度。在今天的X和昨天的Y之间或许有数千种直接的心理联系，也可能只有一个单一的联系。如果只有一个单一的联系的话，照修正的洛克观点，X和Y就不是同一个人了。因为X和Y要是同一个人，则每天都必须有足够的直接心理联系。因为联系性是一个程度问题，我们不能合理地确切界定多少算作足够的。但是如果任何一天直接联系的数目是每个现实的人的生活中每一天所拥有的联系数目的至少一半，我们就能够主张存在着足够的联系性。[6]当具有足够的直接联系时，就有我所称的强联系性。

这关系会是个人同一性的准则吗？如果下述情况为真，关系F是传递性的：如果X与Y是F关系，而且Y与Z是F关系，那么X与Z一定是F关系。个人的同一性是一种传递关系。如果伯纳特与哲学家罗素是同一个人，而罗素与《我为什么不是基督教徒》的作者是同一个人，这个作者与伯纳特一定是同一个人。

强联系性不是一个传递关系。现在的我与昨天的我有强联系，昨天的我与两天前的我有强联系，两天前的我与三天前的我有强联系，等等。但是得不出我现在与二十年前的我有强联系的结论。况且该结论也不为真。在现在的我和二十年前的我之间的直接的心理联系的数目要比大凡成年人任何一天生命中的直接联系的数目要少得多。例如，绝大多数成年人对前一天经验的记忆会有很多，而我对二十年前的任何一天所拥有的经验的记忆却寥寥无几。

我用"时间中个人的同一性准则"意指这同一性所必然涉及的或者必然寓于的。因为同一性是个传递关系，同一性准则也必定是一个传递关系。鉴于强联系是非传递性的，它不能是同一性的准则。而且我也刚刚描述了一个事例，这一点在其中是清清楚楚的。我与二十年前的我自己是同一个人，尽管我现在与那时的我不具有强联系。

尽管洛克观点的辩护者不能诉诸心理上的联系性，但是他能够诉诸心理上的连续性，这是传递性的。他可以诉诸

> 心理准则：（1）有心理的连续性，当且仅当有重叠的强联系链。今天的X与过去某个时刻的Y是同一个人，当且仅当（2）X与Y在心理上是连续的，（3）这个连续性具有那种确切的原因，而且（4）它没有呈现为一种"支线"形式。（5）时间中个人的同一性就寓于事实（2）和（4）的成立之中。

至于（4）按照物理准则的情况如何，将留待后面解释。

心理准则有三个版本。它们在什么是那种正确的原因问题上存在分歧。照狭义的版本，这必须是正常的原因。照广义的版本，这可以是任何可靠的原因。照最广义的版本，这个原因可以是任何的原因。

狭义的心理准则在语词的通常意义上使用这些语词。因此，说我记得有个经验，仅当

（1） 我好像记得拥有一个经验，

（2） 我过去的确有这一经验，而且

（3） 我的明显的记忆从因果上以一种正常的方式有赖于这个过去的经验。

我们需要条件（3）这一点可以通过一个例子来说明。假设我在一次登山运动中摔得失去了意识。在我苏醒过来后，登山的同伴告诉我在我将要掉落时向我大喊了些什么。在以后的某年，当我的记忆不再那么清晰的时候，我可能似乎还记得在我掉落前同伴向我大喊的经验。而且我过去的确拥有这一经验可能是真的。但是尽管条件（1）和（2）得到应对，我们也不应该认为我现在正在记得那个过去的经验。一个确凿的事实是，人们从不能记住他们在失去知觉前那寥寥无几的最后经验。从而我们应该主张，我对听到我的同伴大喊的明显记忆并非是对那过去经验的真实记忆。这明显的记忆不是在因果上以正确的方式依赖于那个过去的经验。我有这明显的记忆仅仅因为我的同伴后来告诉我他所喊的是什么。[7]

同理适用于其他类型的连续性，比如性格的连续性。照狭义的心理准则，即便一个人的性格有巨大的改变，仍然有性格的连续性，如果这些变化具有数个正常原因之一的话。某些性格变化是审慎造成的，另一些则是年龄增大的自然结果，还有一些是对某些经验的自然反应。但是，如果极端的和人所不欲的变化是由不正常的介入而产生的话，诸如对大脑的直接损害，就没有性格的连续性。

尽管是记忆使我们意识到时间中我们自己的连续存在，但是其他各种各样的连续性十分重要。我们可以认为，它们足够重要，足以在即使记忆缺失的情况下提供个人同一性。然后我们将主张（洛克所否定的）即便一个人患有完全的失忆症他仍然继续存在。

除了狭义的版本之外，我还表述了两个广义的心理准则版本。这两个广义的版本展开了数个语词的含义。照"记忆"通常的意思，一个记忆一定有其正常原因。两个广义的心理准则诉诸记忆的广义含义，要么允许任何可靠的原因，要么允许任何原因。同理应用于其他类型

的直接心理联系。为简化我们对这三个准则的讨论起见，我将在最广意义上使用“心理连续性”，它允许这连续性具有任何原因。

如果我们诉诸强调正常原因的狭义版本，心理准则在绝大多数情况下与物理准则相符。记忆的正常原因涉及着大脑的连续存在。而且我们的某些或者所有心理特征有赖于我们大脑中的状态或者事件。一个人的大脑的连续存在至少是心理连续性正常原因的部分。照物理准则，一个人连续存在的条件是：当且仅当（a）连续存在着这个人足量的大脑，以便它仍然保持着是一个活生生的人的大脑，（b）而且在这种物理连续性中没有分支。（a）和（b）据称是一个人在时间中的同一性或者连续存在的充分必要条件。照狭义心理准则，（a）是必要的，但不是充分的。一个人连续存在，当且仅当（c）有心理连续性，（d）该连续性有其正常原因，以及（e）它尚未呈现一个分支形式。（a）作为心理连续性的正常原因的部分是需要的。

再来考虑一下我所想象的那个故事，其中我的大脑和躯体都被摧毁了。扫描仪和复制器制造一个人，这个人具有一个全新的但是确切相似的大脑和躯体，而且这个人在心理上与按下绿色按钮时候的我是连续的。该连续性的原因，尽管不正常，却是可靠的。照物理准则和狭义心理准则两者，我的复制品不会是我。按照两个广义的准则，他会是我。

我将论证我们无须在这三个心理准则的版本之间作出决定。一个部分的类比可以提示为什么。某些人失明是因为眼睛受损所致。现在科学家们正在开发人工眼睛。这些人工眼睛包含一副玻璃的或者塑料的透镜，以及一个微型电脑，该电脑通过视觉神经传送电脉冲，就像一个自然的眼睛通过这种视神经所传送的一样。当这样的人工眼睛更加先进时，或许会赋予已经失明的某个人恢复他一度拥有的那样的视觉经验。他看上去见到的东西会与在他面前事实存在的东西相符合。而且他的视觉经验会在因果关系上，以这种新颖但可靠的方式，依赖于来自他面前的对象的光波。

这个人会看到这些对象吗？如果我们坚持看见必须包括正常的原因，那我们会回答不。但是即使说这个人不能看见，但他所具有的能力好得就像看见一样，既作为知道什么在视力之内的方式，又作为视觉快感的一个源泉。如果我们接受心理准则，我们会作同样的主张。如果心理连续性并不具有其正常原因，它可能并不提供个人同一性。但是即便如此，我们可以主张，它所提供的就如同个人同一性一样。

79.其他观点

我在问时间中个人同一性的准则是什么——这同一性包含什么或者寓于什么。我首先描述了时空的物理连续性，照标准观点，此为物理对象同一性的准则。我然后描述了有关个人同一性的两个观点，即物理的和心理的准则。

有关这些观点存在着一种自然却虚妄的假设。很多人信仰所谓的唯物主义或者物理主义。这是一种这样的观点，即认为不存在纯粹的精神实体、状态或者事件。照物理主义的一个版本，每一精神事件只是某个特定的大脑和神经系统中的一个物理事件。还有其他版本。那些不是唯物主义者的人们要么是二元论者要么是唯心主义者。二元论者认为，精神事件不是物理事件。即使所有的精神事件在因果上依赖于某个大脑中的物理事件，这一点也成立。唯心主义者认为，所有的状态或者事件，当正确理解时，是纯粹精神性的。鉴于这些区别，我们可以假定物理主义者必定接受个人同一性的物理准则。

并非如此。物理主义者会接受心理准则。而且他们会接受允许任何可靠原因或者任何原因的版本。于是他们会认为在简单电子传输中我的复制品会是我。他们会在此处拒斥物理准则。[8]

有关个人同一性的观点并非惟有这些准则。我现在将表述一些其他的观点，它们之所以值得考虑或者出于充分地看似令人信服，或者出于具有足够的支持者。这种描述可能一时难以把握，但却使你对行

将讨论的问题有一个大致的把握。如果这个归纳看起来晦涩难懂或者微不足道，不必担心。

我以一个新区分开始。按照物理准则，随着时间推移的个人同一性仅仅涉及到维持作为一个活生生的人的大脑所需的足量的大脑的物理连续存在。按照心理准则，时间中的个人同一性仅仅涉及各种各样带有正确原因的心理连续性。这两个观点都是还原论者的观点。它们之所以是还原论者的观点，是因为他们主张

（1）时间中一个人的同一性这一事实在于某些更特定事实的成立。

他们可能还主张

（2）这些事实能够在不经预设这个人的同一性的情况下得到描述，或者在不经公开主张这个人生命中的经验为这个人所具有的情况下得到描述，甚至在不经公开主张这个人存在的情况下得到描述。这些事实能够以一种非人格的方式得到描述。

可能看起来主张（2）不会为真。当我们描述统一某人的精神生活的心理连续性的时候，我们必须提及这个人，以及其他许多人，以描述很多思想、愿望、意向和其他精神状态的内容。但是，以这种方式提及这个人并不涉及断言这些精神状态为这个人所具有或者断言这个人存在。这些主张需要进一步的论证，我将把它留待以后来进行。

如果我们对上述两种还原论者的主张都加以拒斥，我们的观点就是非还原论者的观点。

许多非还原论者认为我们是单独存在的实体。照此观点，时间中个人的同一性并不就寓于物理的连续性和／或者心理的连续性。它涉及一个进一步的事实。一个人是个单独存在的实体，有别于他的大脑、躯体以及经验。照这种观点的一个最为人所知的版本，一个人是一个纯粹的精神实体：一个笛卡儿式的自我或者心灵本质。但是我们

可能认为一个人是一个单独存在的物理实体，属于尚未在当代物理学中获得承认的那类。

还有另一种非还原论者的观点。这种观点否定我们是有别于我们的大脑、躯体和经验的一些单独存在的实体。但是这一观点主张，尽管我们不是单独存在的实体，但是个人的同一性是一个进一步的事实，个人的同一性并不就寓于物理的连续性和／或者心理的连续性之中。我称之为进一步事实观点。

我现在将作一些更多的区分。物理准则和心理准则是还原论者观点的一些版本；而且每一准则都有不同的版本。但是一个人的连续存在中所必须涉及的东西少于事实上所涉及的东西。所以，虽然不同准则的信奉者们在有关想象出的事例方面表现出分歧，但是在关于绝大多数真实的人的连续存在中事实上涉及什么方面却表现出一致性。例如，只有当涉及人们开始被电子传输时，他们才会出现分歧。

按照还原论者的观点，每个人的存在不过涉及一个大脑和躯体的存在、某些行动的实施、某些思想的思考和某些经验的发生，等等。展开"事件"一词的通常含义将有所助益。我将用这个词涵盖诸如信念或者愿望的连续存在这样的令人生厌的事件。这种用法使还原论者的观点描述起来更为简单。而且这避免了我深信"精神状态"这些语词会有的一个令人误入歧途的含义。虽然一个状态一定是某个实体的状态，但一个事件并不是如此。鉴于"事件"一词的展开用法，所有还原论者们会接受

（3）一个人的存在就寓于一个大脑和躯体的存在之中，以及一系列互联的物理和精神事件的发生之中。

一些还原论者主张

（4）一个人就是一个特定的大脑和躯体，以及这样的一系列互联的事件。

另一些还原论者主张

（5）一个人是有别于一个大脑和躯体以及这样一系列事件的一个实体。

按照还原论者的观点的这一版本，一个人不纯是一个拥有这些各种各样的组成成分的复合对象。一个人是一个实体，具有一个大脑和躯体，以及具有特定的思想和愿望等等。但是，尽管主张（5）为真，一个人也不是一个单独存在的实体。尽管主张（5）为真，但是主张（3）也为真。

这个版本的还原论可能看起来自相矛盾。主张（3）和（5）可能看起来不一致。考虑一下休谟的类比或许会有所帮助："我再也不能找到比把灵魂比作一个共和国或者一个共同体更为恰当的东西了。"[9]我们绝大多数人在有关国家的问题上都是还原论者。我们会接受下述主张：国家是存在的。鲁里坦尼亚浪漫国（Ruritania）并不存在，但是法国是存在的。尽管国家存在，但是一个国家并非是一个脱离其国民和疆土单独存在的实体。我们会接受

（6）一个国家的存在仅仅涉及那些在其疆土上以某些方式生活在一起的国民的存在。

一些人主张

（7）一个国家就是这些国民和这个疆土。

另一些人主张

（8）一个国家是一个实体，有别于其国民及其疆土。

出于我在附录D中所给出的理由，我们可能认为主张（6）和（8）是不一致的。如果我们相信这一点，我们可能接受在相呼应的

（3）和（5）之间没有不一致性。由此我们可能同意，在主张（3）和（5）中所表达的那个还原论版本是一个一致的观点。如果这个版本是一致的，就像我所认为的，它就是更好的版本。它与我们有关人的实际概念更为接近。但是在下面的绝大多数情况我们可以忽略这两个版本之间的区别。

除了主张（1）和（2）之外，还原论者们可能还主张

（9）尽管个人存在，我们会在不经主张个人存在的情况下对实在给出一个完全的描述。我称该观点为一个能够完全客观描述的观点。

这个观点也可能看起来是自相矛盾的。如果个人存在，而且对什么存在的一种描述没提及个人，这一描述怎么是完全的?

一个还原论者会作出下述回应。假设一个对象有两个名称。就一个既被叫做金星又被称为昏星的行星而言这是真的。在我们对什么存在的描述中，我们会主张金星存在。那么我们的描述就是完全的，即便我们没有主张昏星存在。我们不需要作这一主张，因为通过使用它的别名我们已经主张该对象存在。

当某事实能够用两种方式描述时，同理亦然。一些还原论者接受主张（4），认为个人就是一个大脑和躯体，以及一系列互联的物理和精神事件。如果这就是个人所是的，我们能够在下面的情况下描述这一事实，要么主张

（10）存在一个特定的大脑和躯体，以及互联的物理和精神事件的一个特定系列。

要么主张

（11）一个特定的个人存在。

如果主张（10）和（11）是描述同一事实的两种方法，一个完全的描述无需主张两者。主张（10）就足够了。尽管这个人存在，一个完全的表述无需主张他存在，因为这个事实已经在主张（10）中被表明过了。

其他还原论者接受主张（5），即主张个人有别于他的大脑和躯体，以及他的行动、思想，和其他物理与精神的事件。照还原论的这个版本，主张（10）并非与主张（11）表述同一事实。但是主张（10）可能蕴含着主张（11）。更谨慎起见，鉴于我们对个人这一概念的理解，如果我们知道主张（10）为真，我们应该知道主张（11）为真。这些还原论者们能够说，如果我们对实在的表述或描述或蕴含或使我们能够知道一切存在的事物的存在，我们的描述就是完全的。该主张并不像有关一个完全的描述无需对同一事实给出两个描述的主张那样清晰地为真。但是这一主张看起来像是可取的。如果它获得正当证明，而且还原论者的观点为真，那么，这些还原论者们就能够不经主张个人存在而完全地描述实在。[10]

我的主张就还原论作出了区分，这些区分以这种抽象的形式而言有些难以把握。但是还有其他一些揭示我们在某类事物的观点方面是否是还原论者的途径。如果我们接受一种还原论者的观点，我们应该相信这样的一个事物的同一性可能在一种相当能阐明的方式上是不确定的。如果我们不相信这一点，我们在这类事情的观点上大概是非还原论者。

譬如，考虑一下俱乐部的例子。假设某个俱乐部存在好几年了，它定期举行聚会。后来聚会停止了。若干年之后，该俱乐部的一些成员组成了一个具有同样章程的同名俱乐部。我们问："这些人是重新召集了那同一个俱乐部呢，还是只不过开办了另一个完全相似的俱乐部？"对该问题或许有一个答案。原俱乐部或许就有一个规则解释在经过这样一段时间的不存在之后如何重新召集的问题。也许它有一个不许这么做的规则。但是，假设根本没有这样的一个规则和法律事实来

支持对我们所提问题的两个答案中的任何一个。而且假设有关人员当被问及这个问题时不会给出一个答案。那么对我们的问题就不会有答案。"这是同一个俱乐部"这一主张会是既不真也不假。

尽管我们的问题没有答案，但是可能并没有什么我们不知道的事情。这是因为一个俱乐部的存在并非离开其以某些特定方式共同活动的成员而独立存在。一个俱乐部的连续存在只不过涉及它的成员按照俱乐部的规则举行聚会。如果我们知道有关人们如何举行聚会的所有事实，以及俱乐部的所有规则，我们就知道了一切有待知道的东西。这就是为什么当不能就"这是同一个俱乐部吗？"这一问题作出回答时我们不会感到迷惘的原因。我们之所以不会感到迷惘，是因为即使没有对这个问题给出答案，我们还是能够知道所发生的一切。如果对某个问题而言此点成立，则我称此问题为空洞（empty）问题。

当我们问一个空洞问题时，我们在考虑的只有一个事实或者结果。对我们问题的不同回答只是对该事实或者结果的不同描述而已。这就是为什么不经回答这种空洞问题我们就能够知道一切有待知道的东西的原因。在我所举的例子中，我们能够问"这是同一个俱乐部还是确切相似的另一个俱乐部？"但是在这里并非有两个不同的可能性，其中有一个必定为真。

当一个空洞的问题没有答案时，我们能够决定给它一个答案。我们会决定称后来的俱乐部与原俱乐部是同一个俱乐部。或者我们会决定称它为确切相似的另一个俱乐部。这不是在有关真正发生了什么的问题上的不同观点之间的一个决定。在作出我们的决定之前，我们已经知道所发生的事。我们只是在对同一事件过程的两个不同描述中作一选择。

如果我们在个人同一性上是还原论者，我们应该作出类似的主张。我们能够描述在现在的我和某个未来的个人之间、在物理的和心理的联系上保持递减程度的事例。如果想象自己处于这样一个事例中，我总能够问，"我要死了吗？作为结果的人将是我吗？"照还原论

者的观点，在某些情况下会没有答案。我的上述问题会是空洞的。我将要死去的主张会是既不真也不假。如果我知道有关物理连续性和心理联系性两者的事实的话，我就会知道一切有待知道的东西了。我就会知道一切事情，即使我并不知道我是否将要死去或者是否继续存活许多年。

当这应用于我们自己时，该还原论主张是难以相信的。在这样的一些事例中，某种非同寻常的事情将要发生。但是我们绝大多数人倾向于认为，在任何可以想象出的事例中，"我将要死了吗？"这一问题一定有一个答案。而且我们倾向于相信该答案一定是相当简单的非此即彼的答案，要么是要么否。任何未来的个人一定要么是我要么是其他某个人。我把这些信念称之为我们的同一性一定具有确定性的观点。

下面我将描述两个解释性的主张。第一个回答一个新问题。是什么把一个单一的人同时具有的不同经验统一起来的？在我用打字机打下这个句子的同时，我意识到我手指的运动，我能够看到洒在我书桌上的阳光，而且能够听到风吹树叶的婆娑起舞声。是什么把这些不同的经验统一起来的？一些人主张：它们都是我的经验这一事实。这些是此时一个特定的个人或者经验主体正在具有的经验。一个类似的问题涵盖我的整个一生。是什么把共同构成我的此生的不同经验统一起来的？一些人给出了同一答案。把所有这些经验统一起来的，只是，它们全是我的经验。这些回答我称之为用所有权解释心理统一性的观点。

迄今所描述的那些观点都是关于个人同一性的本质的。我将以一对不是关于这种同一性的本质而是关于这种同一性的重要性的观点来作结束。斟酌一个通常的事例，即使照还原论者的观点的任何一个版本而言，有两个可能的结果。在其中的一个结果中，我将要死去。在另一个结果中我将活许多年。如果这些年会是值得一过的话，那么第二个结果对我来说就是较好的。照绝大多数合理性理论和绝大多数道

德理论，这些结果之间的区别会被判定为很重要。不论我死还是活许多年，它都会具有合理的和道德的重要性。这里被判定为重要的东西是在这些年里是否有一个将是我的人活着。这是一个关于个人同一性的问题。照一种观点，在此类事例中，重要之处总是这一点。我称之为个人同一性是重要的观点。这是自然的观点。

与此相反的观点是个人同一性并非重要。我主张

> 重要的是关系R：有着正确原因的心理的联系性和／或者连续性。

鉴于它更加具有争议性，我追加一个单独的条件

> 在对何为重要的说明中，正确的原因会是任何原因。

正是在想象事例中我们能够最好地决定，重要的是关系R还是个人同一性。一个例子可能是分支事例，其中我的生命短暂地与我的复制品的生命相重叠。假设我与我的复制品是两个不同的人。我将死去，但是我的复制品将再活40年。如果个人同一性是重要的，我应该把我在这里的这样一个前景看作几乎像通常的死亡一样糟的事。但是如果重要的是有着任何原因的关系R，我应该视这种死法为大约像通常的幸存一样好的事。

这些观点之间的分歧并非局限于想象的事例。有关人们实际上所过的所有生命，这两个观点的分歧亦然。在这里分歧较不尖锐，因为照两种观点，所有的或者几乎所有的生命都包含那个重要的关系。照有关个人同一性本质的所有看似可取的观点，个人同一性几乎总是与心理的连续性相符，而大致与心理的联系性相符。但是就像我后面将要论证的，我们相信这些中哪一个是重要的这一点将使事情有很大的不同。如果我们不再认为我们的同一性是重要的，这可能影响到我们的一些情绪，诸如我们对衰老和死亡的态度等等。而且正如我将要论证的，我们可能被引致改变我们有关合理性和道德性的观点。

我现在已经对数个不同的观点给出了一个初步的描述。以这样的抽象方式陈述起来，这个表述可能还不是完全清楚明白。但是现在看似含糊不明的东西，当我讨论这些观点的时候可能变得清晰起来。

这些观点彼此之间有什么联系？我将主张——却正是某些人否定的——这些观点中许多是要么一起成立，要么一起不成立的。如果是这样，将更为容易决定什么是真相。当我们看清这些观点如何联系的时候，我们只有两个选择。我将对要论证的有关观点预先加以陈述。

如果我们不认为我们是单独存在的实体，我们能够可辩护地相信个人同一性是重要的吗？一些作者认为我们能够做到。我将论证我们不能做到。

如果我们不认为我们是单独存在的实体，我们能够可辩护地相信个人同一性并不就寓于物理的和心理的连续性，而是一个进一步的事实吗？我认为我们不能够。

如果我们认为我们的同一性必定是确定的，那我们必定认为我们是单独存在的实体吗？具有第一个信念并不蕴含着具有第二个信念。我们或许既认为我们不是单独存在的实体，又认为对有关个人同一性的任何问题而言必定总是有一个答案，而这个答案必定要么是要么否。有一些学者接受这一观点。但是我将论证这个观点是无可辩解的。只有当我们是单独存在的实体时，我们的同一性必定是确定的这一点才为真。

可能会出现主张我们是单独存在的实体、却否定我们的同一性必定是确定的这种情况。但是鲜有人会把这两个主张结合在一起。

接着假设我们认为心理的统一性是由所有权来解释的。我们认为任何时候一个人的意识的同一性是由这样一个事实解释的，即这个人的不同经验正是为这个人所拥有。而且我们认为一个人整个生命的同一性是由这样一个事实来解释的，即此生中所有的经验是由这个人所拥有的。这些是主张我们是单独存在的实体的那些人们所给出的解

释。如果我们拒斥该主张，我们能够给出这些解释吗？一些学者提出我们能做到。但是我将论证我们不能做到。

我还将为下述结论作论证：

（1）我们不是离开我们的大脑、躯体以及各种各样在物理上和精神上互联的事件而单独存在的实体。我们的存在仅仅涉及我们的大脑和躯体的存在，以及我们行动的实施、我们思想的思考和其他物理的和精神的事件的发生。我们的时间中的同一性不过涉及（a）关系R——有着正确原因的心理联系性和／或者心理连续性，倘若（b）在一个人和两个不同的未来的人之间成立的该关系并不呈现出一个“分支”形式。

（2）我们的同一性总是确定的这一点并非为真。我能够总是问“我将要死去吗？”但是在任何情况下这个问题必定有一个要么是要么否的答案这一点却并不为真。在某些情况下，这会是一个空洞的问题。

（3）有两个有待解释的统一性：任何时候意识的统一性和整个一生的统一性。这两个统一性并非借助主张不同的经验为同一人拥有就能够获得解释的。这些统一性必须借助描述这许多经验之间的关系以及它们与这个人的大脑的关系来加以解释。不经主张这些经验为一个人所拥有，我们就能够指称这些经验，并完全地描述它们之间的关系。

（4）个人同一性并非重要。根本上重要的是有着任何原因的关系R。即使像一个人与另两个人具有R关系的事例中那样，在关系R并不提供个人同一性的情况的时候，该关系也是重要的。另两个关系可能具有些许重要性：物理的连续性和物理的相似性。（在某些人的事例中，诸如那些非常漂亮的人们，物理的相似性可能具有极大的重要性。）

这里是我将如何论证我的结论的简短梗概。我将首先努力回应某些对我有关我们能够以一种非人称的方式描述我们的生命这一主张的诘难。其次我将努力表明，即便我们没有意识到这一点，我们也自然地倾向于认为，我们的同一性必定总是确定的。我们倾向于肯定地认为必定是这样的。接着我将论证，除非我们是单独存在的实体，否则这一自然的信念不能为真。然后我将论证结论（1），即我们不是这样的实体。而且我将论证，因为（1）为真，所以我的其他三个结论亦然。

我们中的大多数人会接受我将予以否定的某些主张。如此，我将论证我们大多数人对我们自己和我们的实际生活具有一种虚妄的观

点。如果我们最终看清该观点是虚妄的，这可能会对我们的生活造成差异。

---

[1]参见，例如，《纸片》（Zettel），G.Anscombe和G.Von.Wright编，G.Anscombe译，布莱克韦尔出版社，1967年，命题350："仿佛是我们的概念包含诸事实的一个脚手架……如果你以另外的方式想象某些事实……那么你就不再能够设想某些概念的应用。"

[2]Quine（1），第490页。

[3]这陈述物体连续存在的一个必要条件。索尔·克里普克在一些讲演中论证说，这个条件不是一个充分条件。鉴于我错过了这些演讲，我不能够讨论这个论证。

[4]照此观点，在像伦敦这样的一个长期以来一直人口稠密的区域生活会是致命的。此处就一度作为许多不同的人最后活着的时候的躯体部分的那许许多多的物质碎屑而言可能为真。这些人们不会全都得以复活，因为不会有足够的此类物质得以重新聚集。一些人持有这个观点的某个版本，避免了这个难题。他们认为，一个得以复活的躯体只需包含原来躯体的一个微粒。

[5]洛克，《人类理解论》第27章，第16节。

[6]这个提法需要扩展，因为对直接联系的数目加以计算的方式有很多。而且一些种类的联系应当被赋予比其他联系更大的重要性。正如我后面提出的，对于那些可以辨别的或者因人而异的联系，应当赋予更大的权重。（例如，所有讲英语的人对如何讲英语分享许多无区别性的记忆。）

[7]我沿袭Martin和Deutscher的说法。

[8]Quinton为这个观点辩护。

[9]Hume（1），第4章第6节，重印载Perry（1）。

[10]还原论提出一些著名的难题。我受到Kripke"命名和必然性"第271页中的这些说法的影响：

尽管英国在1943年与德国作战这个陈述恐怕不能够还原为任何有关诸个体的陈述，然而它在某种意义上并不是"超出"有关诸个人及其历史行动的一个事实。其中关于国家的那些事实并非"超出"关于诸多各人的那些事实的那种意义能够以这样一个观察来表达，那就是，提及关于诸个人的全部事实而忽略关于国家的那些事实的对世界的描述能够是一个对世界的完全的描述，关于国家的那些事实则从之结果产生。同样，恐怕关于物体的事实并非"超出"关于它们的构成分子的事实。那么我们可能问，鉴于对一个就人而言的非现实化的可能情景的描述，英国在那种情景中是否仍然存在……同样，鉴于在桌子T的那些分子的历史中的某些反事实的变迁，有人可能问，T在那种情景中是否会存在，或者，在那种情景中成为桌子的某一束分子是否就构成那张同样的桌子T。在每一情形中，我们借助那些用于别的更加"基本的"殊相的同一性标准来为一些特定的殊相在那些可能的世界中遍寻同一性标准。如果关于国家（或者部落［桌子？］）的陈述不可还原为那些关于更加"基本的"构成成分的陈述，如果在它们的关系之间存在某个"敞开结构"，那么我们就几乎不能够期望给出严格的同一性标准；然而，在具体的情形中我们可能能够回答特定的一束分子是否仍然会构成桌子T，尽管在某些情

形中这个回答可能是不确定的。我认为相似的说法适用于历时的同一性难题……
鉴于那个有关英国之陈述的可能非可还原性，我倾向于弱化克里普克第二句话结尾处的"结果产生"（"follow"）一词。有关个人同一性的主要问题我认为是：这些说法是否不仅适用于国家和桌子而且还适用于人们。

第十一章

我们何以并非自认为的那样

有关个人同一性的不同观点会在有关实际的人们和通常的生活方面产生不同主张。但是当我们考虑某些想象的事例的时候，这些观点之间的差异就更加明显。我将讨论的绝大多数论证部分地诉诸这样的事例。无论科技领域取得的进步可能多么巨大，这些事例中的某一些或许也是不可能发生的。我区分了两类事例。一些事例违反自然规律。我称这些事例为根本上不可能的。另一些则仅仅技术上是不可能的。

如果一些想象的事例将永远不可能出现，这重要吗？这完全取决于我们的问题或者我们要努力表明的是什么。即使在科学领域，考虑根本上不可能的事例也是值得的。其中一个例子就是爱因斯坦的思想实验，他问，如果他能够以光速与光束一同旅行会看到什么。正如这个例子所表明的，我们无须把自己只局限于考虑那些可能的事例。但是我们应该牢记的是，取决于我们的问题，不可能性或许使一些思想试验不相干。

我以对心理准则的一个诘难来开始。

80.心理连续性预先假设了个人同一性吗?

我记得孩提时代曾努力在大西洋拍岸的海浪中保持挺立不动。我与曾经具有这种经验的那个孩童是同一个人。照洛克的观点，使我与那个孩童为同一个人的东西就是我对那个经验的记忆或者说"意识"。

毕晓普·巴特勒认为这是一个"奇妙的错误"。他写道："自明的（是），个人同一性的意识预设了个人同一性，因而不能构成个人同一性，就像在其他事例中预设了真理的知识不能构成真理一样。"[1]

我已经修正了洛克的观点。心理准则不仅诉诸单一记忆而且诉诸记忆的连续性，同时更宽泛地诉诸关系R，后者包括心理连续性的其

他类型。但是这个修正并不回答巴特勒的诘难。

按照某一种解释，诘难会是这样的：“我们只能够记住我们自己的经验这一点是我们的记忆概念的构成部分。记忆的连续性因而预设了个人同一性。你的关系R同理亦然。你主张个人同一性就寓于关系R的成立。如果关系R本身预设了个人同一性，那么这必定是虚妄的。”

为回答这一诘难，我可以界定一个更宽泛的概念——类记忆。在以下条件下我具有对过去经验的确切的类记忆，即如果：（1）我好像记得具有一个经验，（2）某人的确具有这个经验，而且（3）我的明显的记忆从因果上以确当的方式依赖于那个过去的经验。

按照这个定义，通常的记忆是类记忆的一个亚型。它们是对我们自己过去的经验的类记忆。[2]

我们并不类记忆他人的过去经验。但是我们或许开始这么做。长期记忆的原因是记忆痕迹。人们一度认为，这些痕迹或许是定位了的，只涉及到为数不多的脑细胞的变化。现在更加可能的是，一个特定的记忆痕迹涉及更多数目的脑细胞的变化。纵然此为真，假定，神经医生开发出在一个大脑中复制另一个大脑中的记忆痕迹的途径，这或许能够使我们类记忆其他人的过去经验。

考虑一下

威尼斯记忆。珍妮已经同意在她的大脑中复制保罗的一些记忆痕迹。她在手术后的房间里恢复意识之后，拥有全新的一套活生生的明显记忆。她好像记得在大理石铺就的广场上散步，听到鸽子展翅飞翔的声音以及海鸥的鸣叫，看见阳光在碧绿的水面上熠熠生辉。一个明显的记忆十分清晰。她好像记得观看水面那边的一个岛屿，岛上一座白色的帕拉第奥建筑式样的教堂在乌黑的雨云的衬托下十分醒目。

关于这些明显的记忆珍妮应该相信什么？因为她曾经在照片中看到过这个教堂，假定，她知道这是威尼斯的圣乔治教堂。她还知道她从未去过意大利，而保罗经常去威尼斯。既然她知道自己已经接受了保罗某些记忆痕迹的复制，她会有理由假定她可能类记忆保罗在威尼斯的一些经历。

让我们为这个事例加上这样一个细节。珍妮好像记得看到某种非同寻常的东西：乌云中发出一道闪电，这道叉形的闪电既击中了圣乔治教堂的钟楼又击中了正在路过的一艘拖船的红色烟囱。她问保罗是否记得看到这样的一个奇异事件。他看到了，而且还保存了那一期报道这一事件的《小日报》杂志。鉴于所有的这一切，珍妮不应该把她的明显记忆当作一种幻觉来拒绝。她应该得出结论，她具有的看见那道在保罗来说十分确切的闪电是一种类记忆。

由于珍妮的类记忆赋予她有关保罗的经验的知识，她必定大致知道它们是如何导致的。这在通常的记忆事例中是不需要的。除了这一差异之外，类记忆可以提供有关其他人过去生活的一种相似的知识。它们可以从内部提供这些生活像什么的知识。当珍妮好像记得在那个广场四处散步，听到海鸥的鸣叫，看到那座白色的教堂的时候，她知道保罗在威尼斯的那一天的部分情况。

珍妮的明显记忆会在某一方面是错误的。可能有这样的主张："既然珍妮好像记得看到那道闪电，那么她好像记得她自己看到那道闪电。她的明显记忆会确切地告诉她保罗的经验像什么，但是它虚妄地告诉她，正是她具有这一经验。"

在某种意义上此主张可能为真。珍妮的明显记忆可能来自皮科克所称的第一人称表述模式。[3]这样，当她好像记得走过那个广场之际，她或许好像记得看到一个儿童向她跑来。如果这是她好像记得看到的，那么她必定是好像记得她自己看到这个儿童向她跑来。

我们或许否定这些主张。在一个梦境中，我能够好像从外在于我的躯体的角度看到我自己。我或许好像看到我自己向这个视角点跑去。鉴于我好像记得向这个方向跑的正是我自己，这个方向就不能够是向着我自己。我或许说，我好像看到我自己向着看者的视点跑去。而这可以说成是珍妮好像记得的那个儿童所跑的方向。照此说来，珍妮的明显记忆并不包括对她自己的指涉。

尽管我们可以否认珍妮的明显记忆一定好像部分地是关于她自己的，但是无需这样做。即使她的明显记忆是以第一人称模式表述的，珍妮也无需假定如果它们不是幻觉则必定是对她自己的经验的记忆。纵然她好像记得她自己看到那道分叉的闪电，她可以合理地得出结论：她是对保罗的一个经验的类记忆。

珍妮的某些明显记忆显然不是对她自己的经验的记忆。就看到保罗在镜子中刮"她的"胡子的明显记忆而言会是真的。在其他明显记忆的情况下，她或许不得不搞清是她还是保罗具有某个过去的经验。而这有时是不可能的。她可能不得不说，"我的确活生生地好像记得听到那个曲调。但是我不知道是我还是保罗听到的。"当珍妮的明显记忆像这样浮现时，它们在某个方面与我们其他人所具有的记忆的情况有所不同。因为我们并不具有对其他人的过去经验的类记忆，我们的明显记忆并非只以第一人称的模式向我们浮现。它们带着这样的一种信念浮现，即，除非它们是幻觉，它们都是关于我们自己的经验。但是在经验-记忆的事例中，这是一个可区分的信念。如果像珍妮一样，我们具有对其他人过去经验的类记忆，那么这些明显的记忆就会自动地停止与这个信念结合在一起。[4]

现在回到巴特勒对个人同一性的心理准则的诘难。按照这一诘难，记忆的连续性即便部分地也不能使一系列经验之为一个单个人的所有经验，因为这种个人的记忆预设了他的连续的同一性。

按照我在上文所给出的解释，记忆之所以预设了同一性是因为照我们的记忆概念，我们只能够记得我们自己过去的经验。这个诘难现在可以得到回答。我们能够使用类记忆这一更加宽泛的概念。

在我们对我们的修正了的心理准则的陈述中，我们不应该主张：如果我具有对某个过去经验的确切的类记忆，这就使我成为曾经具有该经验的那个人。一个人的精神生活可能包括一些对某些人生命中的经验的类记忆，就像在想象出的珍妮和保罗的事例中的情况一样。我

们的准则忽略了一些这样的类记忆联系。我们取而代之的是诉诸许多这样的联系的重叠链。我们的精神生活寓于一系列各种各样的经验之中。这些包括对早期经验无尽的类记忆。这些类记忆与这些早期经验之间的联系就像一根绳索中的股线一样重叠在一起。如果在每一天直接的类记忆联系的数目至少达到绝大多数人的实际生活中的一半，就存在类记忆的强联系性。强联系的重叠的一股股绳线提供了类记忆的连续性。通过修正洛克的观点，我们主张每一个人的生命的统一性部分地是由这种连续性创造的。我们现在不是在诉诸一个预设了个人同一性的概念。既然类记忆的连续性并不预设个人同一性，它可能是那构成个人同一性的东西的部分。它可能是使我之为现在的我和他日之我为同一个我自己的某种东西的部分。（我说"部分"是因为我们的准则也诉诸其他类型的心理连续性。）

巴特勒的诘难可以以一种不同的方式来解释。他可能本来要表明："在记忆中我们直接意识到我们自己通过在时间上的同一性，而且意识到这是一个独立的、进一步的事实，这一事实不能就寓于物理的和心理的连续性。我们意识到我们每一个人都是一个持久的经验主体，是一个独立存在的实体，这个实体不是我们的大脑或者躯体。而且我们意识到我们的连续存在只是这一经验主体的连续存在。"

我们的记忆告诉我们这一点吗？我们直接意识到这个独立的实体、这个经验主体的存在吗？一些人已经认为，我们不仅在记忆中意识到这一点而且在我们所有的经验中意识到这一点。

81.经验主体

雷德写道：

> 我的个人同一性……意味着我称之为我自己的那个不可分的事物的连续存在。无论这个自我可能是什么，它都是某种进行思维、仔细斟酌、具有决心、付诸行动和遭受苦难的东西。我不是思想，我不是行动，我不是感觉；我是某种进行思考、采取行动和遭受苦难的东西。[5]

在某种意义上，这显然为真。即便还原论者也并不否认人们存在。而且，依照我们对人的概念，人们并非是些思想和行动。他们是些思想者和行动者。我不是一系列经验，而是拥有这些经验的个人。一个还原论者能够承认，在这个意义上，个人是经验的拥有者或者经验主体。鉴于我们谈论问题的方式，此点为真。一个还原论者所否定的是：经验主体是一个与大脑、躯体以及一系列物理和精神事件相分离的独立存在的实体。

在记忆中我们真的直接意识到还原论者所否定的东西吗？我们每一个人都意识到他是一个持久的经验主体，一个并非其大脑和躯体的独立存在的实体吗？我们每一个人都意识到，例如，他是一个笛卡儿式的自我吗？

此点不能得到论证。我并不认为我直接意识到我是这样的一个实体。而且我认为我并非异乎寻常。我认为无人直接意识到这样一个事实。

假设我就算是意识到我是这样的一个实体。仍然存在着对笛卡儿观点的一个诘难。一直有这样一个主张：我不会知道这个实体连续存在。正如洛克和康德两人所论证的，[6]可能存在着一系列在心理上连续的实体。记忆可能就像接力赛中的接力棒一样被从一个人传递给下一个人。所有其他心理特征可能也是如此。鉴于作为结果的心理连续性，我们不会意识到这些实体中的一个已经被另一个所取代。因而我们不能知道这样的实体连续存在。

重新考虑一下分支事例，在该事例中我仍然在地球上这一点是清楚的。可能在某个人看来，他刚好有这两个思想："雪花在飘。所以天气一定很冷。"但是真相可能就是这样。这个人是在火星上的我的复制品。就在我按下那个绿色按钮之前，我想到"雪花在飘"。几分钟以后，在火星上一个相似的舱室中我的复制品突然有了意识。当他有了意识，就有了过着我的生活的明显记忆，而且尤其是他好像记得刚好具有"雪花在飘"的思想。然后他想"所以天气一定很冷"。火星上的我

的复制品现在会处在与我刚好具有这两个思想时的心态确切地相似的一个心态之中。当我的复制品处在这样的一个心态之中的时候，他会认为这些想法两者都被同一个思想者——他自己——所拥有。但这会是虚妄的。我有的是第一个思想，而我的复制品有的只是第二个思想。

这个事例是想象的。但是它好像表明我们不能从我们的经验内容分辨我们是否真的意识到一个独立存在的经验主体的连续存在。我们所具有的充其量是像我的复制品那样的心态。我的复制品虚妄地认为他刚好有两个想法。他并非意识到一个独立存在的实体——这些想法的思想者——的连续存在。他意识到的东西要少得多，即只是他的生命和我的生命之间的心理连续性。同样，当我们具有一系列思想的时候，我们所意识到的充其量也是我们的意识流的心理连续性。一些人主张，我们意识到一些独立存在的经验主体的连续存在。正如洛克和康德所论证的，以及我们的例子好像表明的，这样的意识事实上不能与我们对纯心理连续性的意识区别开来。我们的经验没有赋予我们相信这些实体存在的理由。除非我们有其他认为它们存在的理由，我们应该拒斥这一信念。

这个结论并非像某些学者所写的那样是粗陋的证实主义者的观点。我并非在假定只有我们能够认知的才会是真的。我的评述作了一个不同的假设。我正在讨论有关一类特定事物的存在的一个一般性的主张。这类特定的事物据主张是一个有别于我们的大脑和躯体的独立存在的事物。我主张，如果我们没有理由相信这样的实体的存在，我们应该拒斥这一信念。我并非像实证主义者们那样主张该信念是无意义的。我的主张只是像这样的主张，即，既然我们没有理由相信水中仙女或者独角兽存在，我们就应该拒斥这些信念。[7]

即使我们并非直接意识到这些实体的存在，也有人主张我们能够从我们的经验推论出它们的存在。众所周知，笛卡儿就持这一主张。当他问是否有任何他不能怀疑的事物的时候，他的答案是他不能怀疑

他自己的存在。这是在怀疑这一行动中揭示出来的。而且在假定任何思想必须有一个思想者之外，他还假定一个思想者必定是一个自我或者心灵本质。笛卡儿式的自我是有关一个与大脑和躯体相分离的独立存在的实体的最明显的例子。[8]

利希滕贝格声称，笛卡儿在自认为最明确的方面误入了歧途。他不应该主张一个思想者必定是一个独立存在的实体。他的著名的"我思"并未为这一信念提供正当性证明。他不应该主张"我思故我在"。尽管这是真的，却是误导人的。笛卡儿本该主张"思想：故思生"，或者本该主张，"这是一个思想，故至少一个思想在被思"。[9]

因为我们把思想归因于思想者的活动，我们确实能够断言思想者的存在。但是我们不能从我们经验的内容推演出思想者是一个独立存在的实体。而且正如利希滕贝格所提出的，因为我们不是独立存在的实体，我们可以充分描述我们的思想，而无需声称我们的思想有思想者。我们可以充分描述我们的经验和经验之间的联系，而无需声称它们被一个经验主体所拥有。我们可以给出一种我所称的非人称描述。

正如我已经说过的，一些学者既拒斥还原论者的这个最新的主张，又拒斥笛卡儿的观点。这些学者们并不相信笛卡儿式的自我。而且他们不相信个人是另一类独立存在的实体。他们认为个人的存在就包含着他的大脑和躯体的存在、他的行动的实施和其他各种各样的物理的和精神的事件的发生。但是这些学者们主张，我们不能指称特定的经验或者表述它们之间的联系，除非我们指称某个具有这些经验的个人。依照他们的观点，精神生活的统一性不能以一种非人称的方式获得解释。

斯特劳森讨论了康德提出的对这一观点的一个论证。这个论证主张，我们不能具有对我们周围的世界的认识，除非我们相信自己是些具有我们在时间中的同一性意识的个人。休梅克倡导一种相似的论证。如果这些论证是正确的，它们或许驳斥我有关我们可以用非人称

的方式重新描述我们自己的生活的主张。由于这些论证是处于一种非常抽象的层面，所以我将把它们放到其他更为合适的地方去讨论。[10]

威廉斯讨论了一个对非人称描述的一种较为简单的诘难。[11]这个诘难是针对利希滕贝格的。正如威廉斯所指出的，利希滕贝格提出笛卡儿的"我思"替代物不必完全是非人称的。它不必是"思想：故思生"。它可以是"思想：故我在思"。既然经验主体在这里只是在思想的内容中被提及，这个句子并不把这个思想归属于一个思想者。

然后威廉斯指出，如果数个思想是以此种方式加以表达的，就会需要弄清这些思想是在相同的生活还是在不同的生活中发生的。如果所有这些思想都是以短语"有人认为：……"开始，那么这点就是不清楚的。他一度考虑"（T10）在A处有人认为：……"，但是最后拒斥了这个短语。他继续说：

> ……需要某种更少比喻性的替代物来取代对那个思想之发生陈述中的"在地点A处"——而且得出某种不亚于一个人名或者类似这样的东西将作为一个替代物的结论是自然的，所以T10将让位给
>
> T（11）A认为：……
>
> 在这一点上……引入非人称陈述方式的方案……将会最终崩溃。

威廉斯提出对这一诘难的回答。正如他所写道的，"或许存在着取代那满足有效相对化的比喻性的'地点'的某种替代物，但是尚未走到如此之远，以至于引入一个思考的主体"。存在着许多这样的替代。其中有两个或许是：

> 就包含着以这个句子表达的那个思想之思考的那个特定的生命而言，有人认为……
> 或者
> 就现在直接在因果性上依赖于躯体A的那个特定的生命而言，有人认为……

那么，利希滕贝格会需要以一种非人称的方式解释个人生命的同一性。他会首先修正我们的类记忆的概念。他会主张一个明显的记忆是一个准确的类记忆，如果

（1）明显的记忆是对一个特定过去经验的记忆，

（2）这个经验发生过，而且

（3）该明显的记忆在因果性上以那种正确的方式依赖于该经验。

他会不得不表明那类正确的原因能够以一种并不预设个人同一性的方式得到表述。然后他会诉诸其他类型的心理连续性，诸如一个意向的形成和这一意向得以贯彻的后发的行动之间保持的那种心理连续性。我还得表明一些其他的连续性，以及它们的原因，能够以并不预设个人同一性的方式得到描述。既然它们能够如此被描述，正如我在第89节中所表明的，那么它们也能够以一种非人称的方式得到描述。在描述数不清的思想、愿望和其他经验的内容的时候某些个人是必定被提及的。但是，如威廉斯所指出的，这样的一些表述并非主张这些经验是被某些个人所拥有的。而且，不经作此主张，我们可以表述共同构成一个特定的个人生命的所有精神事件和物理事件之间的那些相互联系。

利希滕贝格对笛卡儿的诘难就这样幸存下来了。我们能够指称和描述不同的思想，以及描述它们之间的关系，而不经把这些思想归属于思想者们。我们事实上的确把思想归属于思想者们。由于我们以这种方式在谈论，笛卡儿会真实地主张，“我思故我在”。但是笛卡儿实际上并未表明一个思想者必定是与大脑、躯体以及各种各样精神事件和物理事件相分离的一个独立存在的实体。

82.非还原论者的观点何以或许本来为真

一些学者主张，笛卡儿式的自我的概念是不可理解的。我对此主张持怀疑态度。而且我认为可能一直存在支持笛卡儿观点的证据。

例如，可能一直存在支持轮回重生信念的证据。一条这样的证据可能是这样的。一个日本妇女或许会声称，她记得她作为一个凯尔特猎人和武士在青铜时代生活过。基于她的明显记忆，她可能作出许多考古学家们可以核对的预测。于是她可能记得有一个铜手镯，状如两条争斗的龙。而且她记得就在她阵亡的那场战斗之前，她在某个特定的巨石旁边把这个手镯埋了下去。考古学家或许现在刚好发现一个手镯埋在这个地点，而且他们的仪器或许显示出此处的土层至少两千年没有动过。这个日本妇女或许还作出了许多其他这样的预测，而且所有的预测都得到了证实。

接着假设有数不胜数的其他事例，其中生活在今天的人们声称曾有过某些过去的生活，而且提供了全部可以得到证实的类似的预言。这适用于当今世界人口中我们绝大多数人。如果有足够的这样的证据的话，而且没有其他方式解释我们绝大多数人如何知晓有关遥远的过去生活的如此细节的话，我们或许不得不承认我们具有有关这些过去生活的准确的类记忆。我们可能不得不承认那个日本妇女具有某种方式，就像对她自己的生活的记忆那般，得悉青铜时代的一个凯尔特武士的生活。

或许接着发现，在凯尔特武士和那个日本女人之间没有物理上的连续性。我们可能因而放弃记忆的载体是大脑这一信念。我们可能不得不假定这些类记忆的原因是某种纯精神的东西。我们可能不得不假定有某个纯精神的实体，这实体过去以某种方式被卷入那个凯尔特武士的生活之中，而现在以某种方式被卷入那个日本妇女的生活之中，而且这实体在把这两个人分割开来的数千年的时间内一直连续存在。笛卡儿式的自我就是这样的实体。如果有轮回重生的充分证据的话，我们可能就有理由相信实际上有这样的一些实体存在。而且我们可能然后有道理地断定这样的实体就是我们每一个人实际上所是的。

这种证据不会直接支持有关笛卡儿式的自我具有笛卡儿的追随者所相信的其他特殊属性的主张。于是它不会表明这些自我的连续存在

要么是一切要么什么也不是。但是可能本来有证据支持这一主张。一个人的大脑或许本来有各种类型或者程度的损坏，一些损坏并未从根本上改变这个人；而其他一些类型或者程度的损坏则好像产生出一个全新的人，与原先的那个人根本没有心理连续性。类似的事情可能本来适用于五花八门的精神疾病。一般来说，我们本来可能会得出这样的结论：这些各种类型的干涉要么对心理连续性毫发无损，要么完全摧毁心理连续性。它本可能证明，发现或者产生具有弱化心理联系性程度的中间事例是不可能的。

我们有确凿的证据相信轮回再生吗？我们有证据相信心理连续性主要不是依赖于大脑的连续性而是依赖于某种其他实体的连续性，而其存在要么毫发无损要么子虚乌有吗？事实上，我们没有上述这类证据。即使我们能够理解笛卡儿式的自我或者心灵本质概念，我们也没有证据相信这样的实体存在。我们也没有证据相信一个个人是其他类型的独立存在的实体。而且我们有相当的证据既相信心理连续性的载体是大脑，又相信心理联系性可以具有任何弱化的程度。[12]

我业已承认非还原论者的观点的最著名的版本——主张我们是些笛卡儿式的自我——可能有些意义。而且我业已提出，如果那些事实是非常不同的，可能本来有充分证据来证明这一观点的正当性。一些相信笛卡儿式的自我的人们并不以这样的方式把那些自我与可观察的事实联系在一起。他们接受洛克和康德所表述的可能性。依照他们的观点，我的笛卡儿式的自我可能突然不再存在，而被另一个自我所取代。这个新自我或许就像在接力赛中一样"继承"了我所有的心理特性。依照这个无特色的笛卡儿式的观点，就当你在阅读本页文字的时候，你可能突然不再存在，而且你的躯体被某个只是与你确切相似的新人所接手。如果此事发生的话，也没有人会注意到任何差异。永远不会有任何证据（不论是公共的还是私人的）表明此事是否发生过，以及如果发生的话是如何发生的。我们因而甚至不能主张这是不太可能发生的。而且有其他一些可能性。按照这一观点，历史或许本来像

往常一样进行，除非我是拿破仑而且拿破仑是我。这不是主张德里克·帕菲特本该是拿破仑。而是主张我是一个笛卡儿式的自我，而拿破仑是另一个笛卡儿式的自我，而且这两个自我可能本来"占据"各自的位置。[13]

当对笛卡儿式的自我的信念以这种方式摆脱了与公众可观察的或者私人可内省的事实的联系的时候，有关它是不可理解的指控变得愈加看似可取的了。笛卡儿的追随者们能否避免他们观点的这个版本并不清楚。并不清楚他们是否能够否定洛克和康德所表述的那个可能性。但是重申我们有拒斥这一观点的充分理由足矣。

83.威廉斯针对心理准则的论证

我已经从两个方面为心理准则作了辩护。我已主张而且部分表明我们能够以一种并不预设个人同一性的方式来表述心理连续性。而且我已主张，按照我们所拥有的证据，这一连续性的载体不是独立于个人的大脑和躯体而存在的一个实体。

下面我将考虑对心理准则的另一个诘难。这是由威廉斯提出的。[14]这一诘难好像要表明，如果某个个人的大脑连续存在而且支持意识，这个个人将继续存在，不论这个个人精神生活的心理连续性中的那些破裂有多么巨大。

这里是该诘难的一个简化版本。考虑一下

威廉斯的例子。我沦为一个无情的神经外科医生的囚徒，他打算通过改变我的大脑来瓦解我的心理连续性。当他给我做手术的时候，我将是意识清醒的，而且处于疼痛中。我因而惧怕将要发生的事情。

外科医生告诉我说，当我处于疼痛中的时候，他将做几件事。他将首先激活一些将引发健忘症的神经元。我将突然丧失直到开始疼痛为止我生命中的一切记忆。这会赋予我对将要发生的事较少惧怕的理由吗？我能够假定当医生拨动这一开关的时候，我的疼痛将突然终止吗？肯定不。疼痛或许如此地占据我的心，以至于我甚至没注意到丧失所有这些记忆。

医生接着告诉我说，当我仍然处于疼痛中的时候，他将随后按下另一个开关，将导致我认为我是拿破仑，而且赋予我关于拿破仑生活的一些明显记忆。我能够假定这将导致我的痛

停止吗？自然，回答还是不。为了支持这一回答，我们可以再次假设我的疼痛将阻止我注意到任何事。我将不会注意到我变得认为我是拿破仑以及获得新的一整套明显的记忆。当这位外科医生按下这第二个开关的时候，在我所意识到的东西中将不会发生什么变化。变化将是纯意向上的。只有如果我的疼痛停止了，所以我能够思考、能够用"拿破仑"这个名字来回答"你是谁"的问题的话，它才变成真的。同样，如果我的疼痛停止了，我然后会开始具有虚幻的明显记忆，诸如视察皇家近卫军或者因1812年灾难性的挫折而流泪的那些记忆等等。如果通过按下第二个开关在我的意向中引发的只会是这样的一些变化，我不会有任何理由期待这导致我的疼痛停止。

那个医生然后告诉我说，在我遭受痛苦折磨期间，他将随后按下第三个开关，那将改变我的性格，以至于变得就像拿破仑的性格一样。再一次在我看来没有理由期待按下这个开关会终止我的疼痛。如果拿破仑的性格与我的性格相比，隐含着更多的坚韧的话，或许充其量带来一些缓解而已。

在这个想象出的事例中，我被告知的一切好像根本没有什么赋予我理由，去期待在受折磨期间我将不再存在。我好像具有同样多的理由惧怕全部的疼痛。这理由好像并不因我不得不惧怕的其他一些事情——失去记忆、发疯、变得像而且自认为是拿破仑——而转移。就像威廉斯所主张的，这个论证好像表明我能够具有惧怕未来的痛的理由，不论在这疼痛之前有什么心理变化。甚至在所有这些变化之后，感到疼痛的还将是我。如果是这样，个人同一性的心理准则是错误的。在这个想象出的事例中，在现在的我与遭受痛苦折磨之后的我之间不会有记忆、性格以及类似事物的连续性。包含在我的连续存在之中的东西从而不能是这样的连续性。[15]

它可能遇到这样的诘难，即，如果我在这个痛苦折磨的整个过程中保持具有意识，那将至少有一种心理连续性。尽管我丧失了我的过去生命的所有记忆，但是我会拥有我遭受痛苦折磨的记忆。尤其是，我会继续具有对最后一些时刻的短期记忆，或者具有有时所称的似是而非的在场。会有这样的记忆的一个重叠链贯穿在我的痛苦感受中。

为了应对这一诘难，我们可以为这个事例添加一个特征。在我失去了我所有的其他记忆之后，我有一刻是无意识的。当我恢复意识的

时候，我没有记忆。随着痛苦折磨的继续，我会有新的记忆。但是在我的无意识的片刻不会存在记忆连续性。

它接着可能遭受这样的诘难，即，我已经以避重就轻的方式描述了这个故事。于是我提出，当我被弄得失去我的记忆的时候，由于我的疼痛我可能没有注意到任何变化。这个描述假定在我的记忆失去之后处于疼痛中的那个人会仍然是我。真相也许是，在这一点上，我会停止存在，而且一个新的个人会开始以我的躯体而存在。

威廉斯会回应说，纵使我的描述假定我会继续存在，这个假定也是完全合理的。反而是心理准则的辩护者必须表明这个假定并非得到正当证明。而且这会是相当难以表明的。难以令人相信的是，如果当我处于剧痛中的时候我被弄得失去意识，这会导致我在痛苦折磨的途中不再存在。而且难以置信的是，在我性格中的变化会具有这一后果。

威廉斯的论证好像驳斥了心理准则。它好像表明为真的观点是物理准则。照此观点，如果某人的大脑和躯体连续存在，而且支持意识，这个人将连续存在，不论这个人的精神生命的心理同一性中的那些裂缝是如何巨大。

84.心理谱系

我现在将修正威廉斯的论证。这样做的原因后面将会论述。

威廉斯讨论的是一个单一的事例，其中在发生了一些变化之后将没有任何心理连续性。我将讨论谱系——或者说一个一系列事例构成的域，谱系中的每一个事例都与其相邻事例非常相似。这些事例囊括了心理联系性的所有可能性。我称之为心理谱系。

在分布在这一谱系最远端的一个事例中，那个外科医生将会导致非常多的开关被同时按动。这会导致在我与作为结果的人之间没有任何心理的联系。这个人会完全像拿破仑一样。

在近端的一些事例中，那个外科医生只会按动为数不多的一些开关。如果他只按动第一个开关，这只会导致我失去一些记忆，而拥有与拿破仑的生命相契合的一些明显记忆。如果他按动前两个开关，我将会失去更多的记忆，而拥有更多的这些新的明显记忆。只有当他按动所有的开关的时候，我才会丧失所有的记忆，而拥有一套完整的拿破仑式的幻觉。

有关我的性格上的那些变化的主张亦然。任何一特定的开关只会导致一个小变化。于是，如果我要像拿破仑，我必定变得脾气更加暴躁，而且必定对看到人们被杀的情景不再感到烦恼不安。按动前两个开关所产生的结果只会是这些。

在包含着众多各不相同的事例的那个论证的修正版本中，我们必须决定我在其中幸存的究竟是哪些事例。在近端的那个事例中，那个医生无所事事。在第二个事例中，我只会丧失一些记忆，拥有一些幻觉，脾气变得更坏而已。显然在这个事例中我会幸存下来。在第三个事例中，变化只是略微更大一些。这点对谱系中任何两个相邻的事例而言都成立。难以令人既相信我会在其中的一个事例中幸存，又相信在下一个事例中我会不再存在。我是否继续存在的问题不能被令人信服地看作是依赖于我是否只会丧失某些记忆、拥有更多的某些幻觉和性格出现些小变化。如果没有这样的小变化会导致我不再存在，我会在所有这些事例中连续存在。我甚至会在谱系最远端的那个事例中连续存在。在这个事例中，在现在的我与作为结果的我之间没有心理联系。

它可能受到这样的诘难：

在这一修正版本中，其论证可疑地与包含累积难题或者说沙堆悖论的那些论证相似。我们被看似无辜的步骤引导到荒谬的结论。大概此处所发生的就是这样。

假设我们主张移去一颗沙粒不可能把一个沙堆改变成某个不是成堆的事物。某人以一堆沙子开始，然后一粒沙一粒沙地移去。我们的主张迫使我们承认在每个变化之后，我们仍然拥有一堆沙，甚至当沙粒的数目变成三、二和一时也是如此。但是我们知道我们得出了错误的结论。因为一粒沙不是一堆沙。

在你诉诸心理谱系中，你主张没有小变化会使你不再存在。但是借助足够多的小变化，那个外科医生会使那个作为结果的人在心理上与你没有任何心理联系。这个论证迫使你断定那个作为结果的人会是你。这个结论可能就像有关沙粒的结论一样虚妄。

为了给威廉斯的论证的这个版本作辩护，我无需解决累积难题。作出下述评述将是足够的。

当考虑堆的时候，我们都相信有些边际事例。两粒沙是一个沙堆呢？还是4粒沙、8粒沙抑或18粒沙是一个沙堆？我们可能不知道如何回答所有这些问题。但是我们并不相信这是无知的结果。我们并不认为这些问题中的每一个必须有一个答案。我们知道堆的概念是个模糊概念，有模糊的边际。当累积论证应用于堆的时候，我们高兴地以一个约定来解决该难题：在有关如何使用"堆"这个词方面的随意决定。我们可能决定我们将不称9粒沙子为一堆沙，但是我们将称10粒或者更多粒数的沙子为一堆沙。然后我们将舍弃该论证的一个前提。依照我们新的更加精确的概念，拿走一粒沙子就可能把一个沙堆变成某种不是一个沙堆的东西。当拿走最后的第10粒沙子的时候，这种转变就发生了。

当它被应用到其他主题时，比方说可感知的颜色时，累积论证就不是如此容易被驳回了。[16]当该论证应用于个人同一性的问题时，这样的驳回似乎并不令人信服。我们绝大多数人都认为在数个方面我们自己的连续存在都不像沙堆的连续存在。

重新考虑一下心理谱系中的事例域。像威廉斯的例子一样，这些事例提供了针对心理准则的一个论证。这个准则是还原论者的观点的一个版本。一个还原论者可能会说：

该论证假定，在这些事例的每个当中，作为结果的那个人要么会是我要么不会是我。并非如此。作为结果的人在最初寥寥无几的那些事例中会是我。在最后那个事例中他不会是我。在处于中间的那许多事例中，两个答案中哪一个都不为真。我可以总是发问，"我要死去了吗？将有某个人过着将是我的生活吗？"但是在这个谱系的中间区域，这个问题没有答案。

> 尽管这个问题没有答案，但是我会确切地知道什么将发生。此处这个问题是空洞的。在每一个这些事例当中，我会知道在什么程度上我与那个作为结果的人之间在心理上会是联系的。而且我会知道哪些特定联系会还是不会成立。如果我知道这些事实的话，我会知道一切事情。我仍然能够问作为结果的那个人是我还是只不过部分像我的某个别人。在某些事例中，这些是两个不同的可能性，其中的一个必定是真的。但是在这些事例中，它们不是两个不同的可能性。它们只不过是对事件的同一过程的两种不同表述。

这些评述与应用于沙堆时我们所接受的评述是可类比的。我们并不认为沙子的任何聚集必定要么是沙堆要么不是沙堆。我们知道有边际事例，其中不存在对"这仍然是一个沙堆吗？"这一问题的明确答案。但是我们并不认为在这些事例中，必定有一个要么是要么否的答案。我们认为在这些事例中，这是一个空洞的问题。甚至不经回答这个问题，我们就知道一切。

正如威廉斯所主张的，当这样的评述应用于我们自己的存在时，好像是令人难以置信的。假设在这个谱系的中间区域我将经受一个手术。我知道那个作为结果的人会处于极度痛苦中。如果我不知道我是否将是处于极度痛苦中的那个人，而且我甚至不知道自己是否将活着，我如何能够相信我确实确切地知道将要发生的是什么？我并不知道对于这些最重要的问题的答案。难以相信这些问题是空洞的问题。

我们绝大多数人相信我们并不像沙堆，因为我们自己的同一性必定是确定的。我们相信即使在这样的一些"边际事例"中，"我将要死去吗？"这个问题必定有一个答案。而且正如威廉斯所主张的，我们相信答案必定是两种相当简单的答案之一，即是或否。如果某人将活着，而且将遭受极度的痛苦，这个人要么是我要么不是我。其中的一个必定为真。而且我们不能使第三种选择有意义，诸如处于极度痛苦中的人部分地是我。我能够想象我只是部分地处于极度痛苦中，因为我在有意识和无意识的状态之间漂移。但是如果某人完全意识到该极度痛苦，这个人不可能部分地是我。

还原论者的观点会对威廉斯的论证提供一个答案。当威廉斯给出这个论证的版本时，他拒斥这一观点。相反他断定，如果我的大脑连

续存在而且是一个活生生的人的大脑，我将是那个人。即便在现在的我自己和后来的我自己之间没有心理的联系，也是如此。在提出他的论证之后，威廉斯写道，他的结论可能"大概"是错的，"但是我们需要被表明错在哪里"。[17]

85.物理谱系

一个诘难是相似的论证应用于物理连续性。考虑一下另一个可能的事例域：物理谱系。这些事例囊括物理连续性的所有不同的可能性。

在这个谱系的那个近端事例中，后来会有一个与现在的我在心理和物理上完全连续的一个人。在远端那个事例中，后来会有一个与现在的我在心理上而不是在物理上连续的一个人。远端的事例就像电子传输事例一样。近端事例是连续存在的正常事例。

在一个靠近近端事例的一个事例中，科学家们会用确切的复制物置换我大脑和躯体中1%的细胞。在居于谱系中间的那个事例中，他们会置换50%的细胞。在靠近远端事例的一个事例中，他们会置换99%的细胞，只留下我原先大脑和躯体的1%。在远端事例中，"置换"将包括我的大脑和躯体的完全毁灭，以及用新的有机物质创造出一个复制品的我。

在最后这个事例中重要的并不只是我的复制品的大脑和躯体会完全由新物质构成。正如我解释过的，以一种并不毁灭我的大脑和躯体的方式来进行会成为真的。如果在我的躯体的物质中有一系列漫长的微小变化，期间我的大脑和躯体连续存在而且功能正常，它会变成真的。这会像那艘在经过50年的小补小修之后变得完全由新木料构成的船一样。在这两个事例中，构件的同一性中的完全变化并未消除物理的连续性。在物理谱系远端事例中情况则不同。这里没有物理连续性，因为我的大脑和躯体完全毁灭了，而且只是后来科学家们用新物质创造出我的复制品。

在这个领域中的那些为数不多的最初事例中的情况，现在据信在技术上是有可能的。部分大脑组织已经成功地从一个哺乳动物的大脑移植到另一个哺乳动物的大脑。而且所移植的会是在所有个体中都十分相似的一个大脑部分。这会使外科医生能够为大脑的某些受损部分提供功能正常的置换物。这些实际上的移植证明比那些人们较多熟悉的肾脏或者心脏移植更加容易，因为大脑好像并不像躯体排斥被移植的器官那样"排斥"被移植的大脑组织。[18]尽管这个领域中最初的几个事例甚至现在就是可能的，但是其中的绝大多数事例将仍然是不可能的。但是这种不可能性将是纯技术上的。因为我用这些事例只是要揭示我们所相信的是什么，所以这个不可能性无关紧要。

假设我们认为在这个谱系的远端，我的复制品不是我。他只是非常像我的某个人。在这个谱系的近端，没有置换，作为结果的人是我。如果将发生的是某个居间事例，我们应该期待什么？如果科学家们只置换1%，我会不再存在吗？这并不会发生，因为我并不需要我的大脑和躯体中的一切。如果他们置换10%、 30%、 60%或者90%，情况又当如何？

这个事例域对物理准则——还原论者的观点的另一个版本——形成了挑战。设想一下你要经受这样一个手术。你或许试图相信还原论的这个版本。你可能对自己说：

> 在这个领域中的任何中心事例中，"我将要死去吗？"这个问题没有答案。但是我刚好知道将要发生什么。我的一定百分比的大脑和躯体将用所存在的细胞确切复制物来置换。作为结果的人与现在的我在心理上是连续的。这是有待知道的一切。我并不知道作为结果的人将是我，还是只与我确切相似的某个别人。这里它不是一个必须具有答案的一个问题。它并非表述两个不同的可能性，而其中的一个必定为真。这里它是一个空洞的问题。作为结果的人是我与是某个别人之间没有真正的区别。这就是为什么说尽管我并不知道我是否将要死去，但是我知道一切的原因。

我认为就那些接受物理准则的人们而言，这是对这个事例域的正确反应。但是我们绝大多数人还不会接受这样的一些主张。

如果我并未接受还原论者的观点，而且继续相信我们的同一性必定是确定的，对这些事例我们应该作何主张？如果我们继续假定我的复制品不是我，我们就被迫得出下面的结论：必定有某个如此这般的关键比例，以至于如果外科医生置换的比例小于它，苏醒过来的将是我；如果大于它，则苏醒过来的将不是我，而只是像我的某个别人。我们或许提出这个结论的一个变化形式。大概我的大脑有某个关键部分，以至于如果外科医生不置换这一部分，那个作为结果的人将是我，但是如果他们置换了这一部分，作为结果的人将是某个别人。但是这没有什么两样。如果他们置换我的大脑的这个关键部分的不同的比例，又当如何？我们再次被迫面对必定有某个比例的观点。

这样的一种观点并非不连贯。但是难以置信。而且其他某事为真，使之更加难以置信。我们通过完成这个想象出的事例域中的某些事例，不会发现这个关键的比例是多少。我或许说，“试着置换我的大脑和躯体50%的细胞，我将告诉你发生什么”。但是我们预先知道，在每一个事例中，作为结果的人会倾向于相信他是我。而这不会表明他是我。完成这样的一些事例不会为我们的问题提供答案。

这些评述假定一个人的所有心理特征依赖于他的大脑和神经系统中细胞的状态。我假定我的一个有机复制品会在心理上确切地像我。如果我们拒斥这个假定，我们对这个想象出的事例域的反应方式就会不同。在下一节中我回答这一反应。

如果我的假定是对的，而且这些作为结果的人中每一个都会确切地像我，有关这个事例域我们应该相信什么？我们有三个选择：

（1）我们会接受上述的还原论的回应。

（2）我们会认为在两个事例之间存在着鲜明的边际。如果那些外科医生只置换某些细胞，那作为结果的人会是我。相反如果他们置换更多一些的细胞，那作为结果的人就不会是我，只会是确切像我而已。在这个事例域的某处必定有一个这样的鲜明边际，即便我们决不会发现这个边际在什么地方。

（3）我们会认为，在所有这些事例中，那作为结果的人会是我。

就这三个结论而言，在绝大多数人看来（3）是最令人信服的。如果我们接受（3），我们认为心理连续性提供个人同一性。即使当这连续性并不具有其正常原因——某个特定大脑的连续存在——的时候，我们也如此认为。

当我们前面考虑心理谱系的时候，威廉斯的论证好像要表明心理连续性对个人同一性而言并不是必需的。物理连续性就足够充分了。当我们考虑物理谱系的时候，一个类似的论证好像要表明物理连续性对个人同一性而言不是必需的。心理连续性就足够充分了。

我们会对这两个结论都加接受。我们会主张两者中的任何一种连续性都提供个人同一性。尽管这个混血观点是连贯的，但是却面临一些严峻的诘难。如果我们不是联合我们的这两个结论，而是联合这些结论的两个论证，一个诘难就此产生。

86.联合谱系

考虑一下另一个可能的事例域。这些事例涉及物理的和心理的联系性两者在程度上的各种可能的变化。此为联合谱系。

在这个谱系的近端是个正常事例，其中一个未来的个人会与现在的我在物理上和心理上完全连续。这个人会是我，就像在我的实际生命中明天醒来的将是我一样。在这个谱系的远端，作为结果的人与现在的我既没有物理上的连续性又没有心理上的连续性。在这个事例中，科学家们将毁灭我的大脑和躯体，然后用新的有机物创造出某个别人的一个完美的复制品。我们假设这个人不是拿破仑而是格丽泰·嘉宝[19]。我们可以假设当嘉宝30岁的时候，一组科学家记录下了她的大脑和躯体的所有细胞的状态。

在这个谱系的最初一个事例中，在谱系的近端，科学家什么都不会做。在第二个事例中，我的大脑和躯体中的一些细胞被置换。新细胞不是确切的复制物。结果，在我和要苏醒过来的人之间会有微小的心理联系性。这个人不会具有我的所有记忆，而且他的性格会在某个

方面与我不同。他会具有格丽泰·嘉宝生命中的某些明显的记忆，并具有嘉宝的某些性格。与我不同，他会喜爱表演。他的躯体也会在某个方面较少像我的躯体，而较多地像嘉宝的躯体。他的双眼会更像嘉宝的那双明眸。沿着谱系进一步深入，我的更大比例的细胞被置换，再次是以不相同的细胞置换。作为结果的人与我在心理上的联系更少，与30岁时的嘉宝在心理上的联系更多。而且在这个人的躯体中会有类似的变化。在靠近谱系远端的事例中，我的绝大多数的细胞会被不相同的细胞所置换。醒过来的人只有我原初大脑和躯体中的为数不多的细胞，而且在她和我之间只有为数不多的心理联系。她会有一些与我的过去相符的明显记忆，以及一些习惯和愿望。但是在任何其他的方面，她会既在物理上又在心理上就像格丽泰·嘉宝一样。

我相信这些事例为还原论者的观点提供了一个强论据。这个论据再次假定我们的心理特征依赖于我们大脑的状态。假设心理连续性的原因不是大脑的连续存在，而是某个像笛卡儿的自我那样的独立存在的实体。那么我们会主张，如果我们实施这些手术，那些结果不会像我所表述的那样。我们将发现，如果我们置换某人相当多的大脑细胞，即使用不相同的细胞加以置换，作为结果的人将非常像原先的那个人。但是将会有某个关键的百分比或者大脑的某个关键部分，其置换将会完全摧毁心理连续性。在这个事例域中的一个事例中，连续性的承载者将会要么不再存在，要么不再与这个大脑互动。作为结果的人将会在心理上完全不像原先的那个人。

如果我们有理由相信这个观点，它将会为我们的论据提供一个答案。在这个事例域中，将会存在一个鲜明的边际。而且能够发现这个边际会与个人同一性中所发生的一个完全变化相对应。这个观点也会解释一些细胞的置换如何可能完全摧毁心理上的连续性。而且这个观点既可被应用于心理谱系又可被应用于物理谱系。我们可以主张在这两个谱系中，结果事实上不是我们所假定的。

除了靠近联合光谱近端的那些事例之外，其中别的那些事例在技术上是不可能的，而且很可能会持续不可能。因此，我们不能够直接发现那些结果会是我们所假定的那样，还是就像所表述的那样。那些结果将是什么依赖于某人大脑的状态与这个人的精神生活之间的关系是什么。我们有证据相信心理连续性主要不是依赖于大脑的连续性，而是依赖于要么毫发无损要么根本不存在的某个别的实体的连续性吗？事实上我们并不具有我上面所表述的此类证据。而且我们有很多理由既认为心理的连续性的承载者是大脑，又认为心理的联系性能保持到任何减弱的程度。

既然我们的心理特征依赖于我们大脑的状态，这些想象出来的事例只是在技术上是不可能的。如果我们能实施这些手术的话，那些结果将会是我已所表述的。关于这个联合谱系中的不同事例我们应该相信什么？其中我将会连续存在下去的那些事例是什么？

就像以前，我们不可能通过在我和他人身上实施此类想象出的手术来寻求答案。我们已经知道，沿着谱系到某处会有首个事例，其中作为结果的人会相信他或者她不是我。而且我们没有理由信赖这个信念。在此类事例中，某人是谁是不可能用他认为他是谁来表明的。既然实验不能有所帮助，我们现在必须努力决定关于这些事例我们应当相信什么。

在考虑最初这两个谱系的时候，我们有三个选择：（1）接受还原论者的一个回应，（2）相信必定有一个边际，以及（3）相信作为结果的人在任何情况下都是我。在这三者中，第三个看来最令人信服。

在考虑联合光谱的时候，我们不能接受这个结论。在远端的一个事例中，科学家们毁灭了我的大脑和躯体，然后用新的有机物创造出格丽泰·嘉宝的一个复制品。在我和这个作为结果的人之间不会有任何联系。在这个事例中，作为结果的人不会是我这一点并不更加清晰。我们不得不在其他两个选择中作出选择。

我们或许继续认为我们的同一性必定是确定的。我们或许继续认为，对于"作为结果的人会是我吗？"这一问题，必定总是有一个答案，而且是相当简单的是或否。那么我们会被迫接受下述主张：

> 在这个谱系的某处，有一个鲜明的边际。必定有被置换的某个关键的细胞群，以及心理变化的某个关键程度，正是这会造成所有的差别。如果外科医生置换略少于这些细胞的细胞量，以及产生小于关键程度的心理变化，苏醒过来的将是我。如果他们置换超额的细胞，以及产生过量的心理变化，我将不再存在，苏醒过来的人将是某个别人。在这个谱系的某处必定有这样的一对事例，纵然永远不会有任何这些事例处于何处的证据。

这些主张难以置信。难以相信（1）生与死的差异就能寓于上述任何非常细小的差异。我们倾向于认为在某个未来的人是我与他是某个别人之间总有一个差异。而且我们倾向于认为这是一个深层的差异。但是在谱系中相邻事例之间这种差异是微不足道的。因而难以相信在这些事例中的一个事例中，作为结果的人会肯定地是我，而在接下来的一个事例中他会肯定地是某个别人。

也难以令人相信的是（2）在谱系的某处必定有这样一个鲜明的边际，尽管我们从未有过这个边际在何处的证据。一些人会主张如果从未有过这样的证据，那么，主张某处必定有这样的一个边际就没有意义。

即便（2）有意义，主张把（1）和（2）放到一起看，也是极为不可取的。我认为它们甚至比另一种惟一可能的结论——还原论者的观点——更加不可取。因而我们现在应该断定还原论者的观点为真。按照这个观点，在联合谱系的那些主要事例中，作为结果的人是否是我这个问题会是一个空洞的问题。这个谱系，正如我所主张的，为这个观点提供了一个强论据。

有一些人认为我们的同一性必定是确定的，尽管他们并不认为我们是与我们的大脑、躯体以及我们的经验相分离的独立存在的实体。我认为这个观点是不能得到辩护的。什么解释个人同一性总是确定的这一所谓的事实？答案必定是个人同一性的真正准则涵盖每个事例。

真正准则必定在联合谱系的某处划下一个鲜明的边际。但是，如果我们不是独立存在的实体，怎么会有这样的一个边际？什么能使作为结果的人在一个事例中会是我而在下一个事例中不会是我这一点为真？差异寓于何处？

另一些人认为，尽管我们不是独立存在的实体，但是个人同一性是一个进一步的事实。这些人认为个人同一性并不就寓于物理的和心理的连续性的不同类型之中。这是另一个我认为是不可辩护的观点。如果我们不是独立存在的实体，这个进一步的事实又寓于何处？什么能使这个事实在这个事例域中的一些事例中或是成立或是不成立？

我认为这个谱系表明，某些观点必须放到一起来共同把握。我们无可辩驳地相信我们的同一性涉及到一个进一步的事实，除非我们还认为我们是与我们的大脑和躯体相分离的独立存在的实体。而且我们无可辩驳地相信我们的同一性必定是确定的，除非我们认为这些独立存在的实体的存在必定是要么全有要么全无。

一些人认为一切事物的同一性必定总是确定的。这些人接受了没有同一性就没有实体这一学说的一个严格形式。这是这样的一种主张，即，我们不能指称一个特定的对象，或者命名这个对象，除非我们有关这个对象的同一性的准则在任何可想象的情况下都产生一个明确的答案。按照这个观点，我们常常在没有某个对象的同一性准则的时候错误地认为我们是在指称某个对象，因为没有这样的准则就没有这样的对象。于是会有这样的主张：我们绝大多数人错误地认为专名“法国”指称一个国家。照这个观点，国家是不能被指称的，因为它们并不存在。对国家而言，没有符合所需标准的同一性的准则——告诉我们在一切可以想象的情况下某个国家是否已经连续存在。那些坚持这个观点的人们可能认为，个人并不存在这一点并不同样为真。如果这个观点为真，而且个人的确存在，个人同一性的准则在所有情况下必须产生一个明确的答案。

这个观点无需包含个人是独立存在的实体这一信念。这可能使这个观点更加可取。但是，如果我们坚持这个观点，我们必须再次相信个人同一性的真正准则在联合谱系的某处划了一个相当不可知的鲜明边际。正如我已经主张的，如果个人的同一性并不涉及到一个进一步的事实，那么认为能够有这样一个边际是很难的。这甚至比还原论者的观点更加不可取。

还有一些学者以另一种方式主张我们的同一性必定是确定的。按照这个观点，如果存在着我们不能回答有关某个对象的同一性的问题的事例，那么我们的信念就前后不一致。我认为有这样的一些事例，而且在这样的事例中某个对象的同一性是不确定的。我主张，在这样的一个事例中，“这是我以前拥有的同一个对象”这一陈述会是既不真也不假。有人论证说这个主张是前后不一致的。[20]我相信这个论证业已得到了回答。[21]但是假设这个论证是正确的。这蕴含着下面的情况。当我们发现并不为我们所相信的同一性的某个准则所涵盖的事例的时候，我们应该通过延展这个准则来修正我们的信念。如果我们持有这个观点，我们并不认为个人同一性的真正准则必须在联合谱系的某处划下某个鲜明的边际。我们宁可认为，为了避免前后不一致性，我们应该划下这样一个界限。

这个观点与还原论者的观点几乎没有区别。如果我们不划这样一条线，我们并不认为它固有地具有合理的或者道德的重要性。我们必须在这个谱系上选取某个点，到这个点之前我们称那作为结果的人是我，超过了这个点我们将称之为某个别人。我们的这个点的选择将不得不是随意的。我们必须在两个相邻事例之间划下这条线，尽管在这两者本身之间的差异是微不足道的。如果这是我们所做的，这不应该影响我们对待这两个事例的态度。把第一个事例当作像通常的幸存一样好，而把第二个事例当作像通常的死亡一样糟，这一点对我来说明显会是非理性的。当我考虑这个事例域的时候，我自然地发问，“作为结果的那个人将是我吗？”通过划下我们的界限，我们已经选择给予这

个问题一个答案。但是，由于我们的选择是随意的，它不能正当证明任何有关重要的是什么的主张。如果这是我们如何回答有关我的同一性问题的方式，我们就业已使之为真：在这个事例域中，个人同一性不是重要的。而且这是还原论者的观点中最重要的主张。我们的观点只是在细微之处与之有差异。还原论者们主张，在一些事例中有关个人的同一性问题是不确定的。我们追加了这样一个主张，即在这样的一些事例中我们应该给予这些问题以答案，即便我们不得不这样做的方式是随意的，而且它剥夺了我们的回答的任何重要性。我把这个观点当作还原论的一个版本，这个头脑清醒的版本废除了带有令人不感兴趣的约定性定义的不确定性。由于差异是如此之微小，我将忽略持此观点的这个版本。

照物理主义的最简单版本，每一个精神事件都是某个大脑中的事件。我在上文评述过，我们能既是物理主义者，又接受个人同一性的心理准则。我应该加上还原论者不必是物理论者这一点。如果我们不是物理论者，我们要么是相信精神事件不同于物理事件的二元论者，要么是相信所有的事件是纯粹的精神事件的唯心主义者。如果我们相信我们是些笛卡儿的自我，那我们就是相信二元论的一种形式。但是二元论者能够是关于个人同一性的还原论者。我们能够相信精神事件不同于物理事件，而且相信一个人的生命的统一性就寓于所有的精神的和物理的事件之间的各种各样的联系中，所有的精神和物理事件共同构成这个生命。这是还原论者的观点的二元论版本。

我将论证，如果我们是还原论者，我们不应该尝试在个人同一性的不同准则之间作出决定。一个理由是个人同一性并不重要。在我为这个结论论证之前，我将进一步解释一个还原论者主张什么。而且既然我们绝大多数人强烈地倾向于拒斥这些主张，考虑联合谱系可能不足以改变我们的观点。因而我将在下一章中提出还原论者的观点的其他一些论证。

还原论者们承认在量的同一性和确切相似之间存在区别。在某些事例中，在某个人是我和他是某个只是确切像我的人之间会具有真正的差异。许多人假定必定总是有这样一种区别。

在那个有关国家或者俱乐部的事例中，这样的一种假定是虚妄的。两个俱乐部可以同时存在，除了其会员外确切相似。如果我是其中一个俱乐部的成员，而你也主张是会员之一，我或许问，“你是我同一个俱乐部的会员吗？还是另一个确切相似的俱乐部的会员？”这不是一个空洞问题，因为它表述两个不同的可能性。但是，尽管在两个俱乐部共存这个事例中有两种可能性，但是我们在讨论某个现存的俱乐部和一个过去的俱乐部之间的关系时可能没有两个这样的可能性。我在第79节所述的那个事例中，就没有两个可能性。在那个事例中，没有什么会正当证明我们有那同一个俱乐部这一主张还是我们有一个只是确切相似的一个新俱乐部这一主张。在该事例中，这些不会是两个不同的可能性。

同样，在一些事例中，在某个人是我和他是十分像我的某个别人之间存在着真正的差异。在分支事例——那个扫描仪并不摧毁我的大脑和躯体的电子传输版本——中可能就是如此。在分支事例中，我的生命与火星上我的复制品的生命相重叠。鉴于这个重叠，我们可能断定我们是两个不同的人——我们是质上同一但不是量上同一。如果我是地球上的那个人，而且火星上我的复制品现在存在，是我将感到疼痛还是我的复制品感到疼痛就产生了不同。这是一个在将发生的事情中的一个真实差异。

如果我们回到在我的生命和我的复制品的生命之间没有重叠的那个简单电子传输的事例，情况则不同。在这个事例中我们可以说我的复制品将是我，或者相反，我们可以说他将只是与我确切相似的某个别人。但是我们不应该把这些当作有关将要发生什么的相互竞逐的假设。这些要成为相互竞逐的假设，我的连续存在必须涉及到一个进一步的事实。如果我的连续存在只是牵涉到物理的和心理的连续性，我

们就知道在这个事例中发生什么。将有某个未来的人，他在物理上将确切地像我，而在心理上将与我完全连续。这个心理连续性将具有一个可靠的原因，将我的图像传输。但是这个连续性将不具有其正常的原因，因为这个未来的人在物理上将不与我相连续。这是对有关事实的一个完全表述。没有我们忽视了的进一步的事实。如果个人的同一性并不涉及到一个进一步的事实，那么我们不应该相信这里有两个不同的可能性：要么我的复制品将是我，要么将是某个只是像我的别人。什么会造成这些不同的可能性？差异能会寓于何处？

一些非还原论者会同意，在这个事例中，不存在两个可能性。这些人认为，在电子传输这个事例中我的复制品不会是我。我将在后面讨论这个结论的一个可取的论证。如果我们说我的复制品是我会是错误的，我刚作的评述也适用于物理谱系中的那些主要事例。我的复制品或许具有我的大脑和躯体中存在的细胞的1／4、 1／2或者3／4。在这些情况下，没有下述两个不同的可能性：我的复制品要么是我，要么是某个只是像我的别人。这些只是对同一结果的不同表述。

如果我们相信在某个人是我与某个人是别人之间总是存在着一个真实的区别，我们必须相信这个差异来自这个事例域中的某处。必定有一个鲜明的边际，尽管我们从不知道这个边际在哪里。正如我已经主张的，这个信念甚至比还原论者的观点都更不可取。

在有关俱乐部的那个事例中，尽管在量的同一性和确切的相似性之间有时有差异，但是有时没有差异。“是同一个抑或只是确切相似？”这一问题有时是空洞的。就人的事例而言情况也会是如此。要么在物理谱系的末端会如此，要么在中央处会如此。

难以相信这会是真的。当我想象我自己将要按下绿色按钮的时候，难以相信不存在我是将要死去还是将在火星上苏醒过来这么一个真正的问题。但是正如我已经论证的，这个信念不能得到正当证明，除非个人同一性涉及到一个进一步的事实。而且不会有这样一个事实，除非我是一个与我的大脑和躯体相分离的独立实体。笛卡儿的自

我就是这样的一个实体。正如我已经主张的，没有有利于这个观点的证据，而且有许多证据不利于它。

---

[1]Butler，第100页。

[2]我沿袭Shoemaker（1）。

[3]见Peacocke（2）。

[4]Evans（2），第246页正确地批评了Parfit（1）中对于记忆的误描述。埃文斯还反对“这样一种官能的可能性，它既与给予诸主体过去知识中的记忆想象，又因为记忆状态的内容绝不侵蚀关于谁的过去这个问题而与之不像”（第248页）。这与我主张类记忆可以提供一种了解他人过去经验的途径相冲突。埃文斯的论证是这样的：
假设我们用外科手术的方式“转移记忆”，从主体S'的大脑到主体S的大脑；而且假设S并不知道这已经发生。S当然将以通常的方式对他的过去作出判断。但是假设他发现他既不曾是F，也不曾是G，…… ——总而言之，他的记忆不能够被当作对于他的过去的精确记录来依赖。让我们美好地假设，然后他退而作出一般过去时态的判断：“某人是F，而且是G……”这些判断不可能会构成知识。在提出它们的时候要达到可理解，S甚至也会不得不把实际上本会发生什么，或者某种与之非常相似的东西，作为假说提供出来；但他不可能会被说成知道它是真的。它会是一个纯粹的猜测。从而他不会被说成知道基于它之上的任何东西……我们因而没有发现什么令人信服的理由来放弃这样一个观点，就是，我们关于我们自己之观念并不允许在借助记忆的运作知道某人看到一棵树燃烧与知道是某人自己看到一棵树燃烧之间存有裂隙（第244—245页）。
我接受埃文斯的主张，就是他所设想的人S不会在他对S'的经验的准记忆中具有任何关于这些过去经验的知识。正如他所写到的：“一个真信念在因果上有赖于使之为真的那些事实这一点对于知识而言并不是充分的。”这些准记忆之所以并不给予S知识，是因为S对于它们的原因一无所知。他并未被告知说，那些外科医生已经在他的大脑中复制了S'大脑中的记忆痕迹。对于珍妮而言情况有所不同。她知道她对于保罗的经验的类记忆如何在因果上有赖于那些经验。埃文斯的论证并未表明这些准记忆不能够给予珍妮知识。它并未从根本上削弱我有关准记忆可以提供一种了解他人过去经验的途径的主张。正如威金斯所提及的，我们能够想象这种知识是一个自然现象。子女在他们自己成胎的那一刻之前就具有对于双亲生活中的经验的一些类记忆本来或许是真的。如果这样的准记忆呈现出这个正规的、受限的形式，那么它们就会提供知识。（Wiggins（4），第145页，这个提法大概归功于H·伊希果罗。）

[5]Reid，重印载Perry（1），第109页。

[6]参见Locke，第27章，第13节，重印载Perry（1），第101页；以及Kant，第342页，特别是注释a。这个例子我引自Wachsberg。

[7]非还原论者还有数个其他的论证，对此在一个更长的讨论中我本来会需要努力予以回答。除了斯温伯恩之外，请特别参见Lewis（4）、（5）和（6），以及Madell。我希望另寻他处对这些论证加以讨论。

[8]Descarts，第101页："我注意到，当我如是切望认为所有事物都假的时候，如此想的'我'是某物这一点是绝对必需的，而且'我思故我在'这个真理是如此确定和如此确实，以至于提出来的所有最放肆的假定……不能够动摇它，我得出结论，我会毫不迟疑地把它当作我正在寻求的哲学第一原理来接受。"

[9]Lichtenberg，第412页。（原文是：'Es denkt，sollte man sagen，so wie man sagt：es blitzt'.）要讲清该箴言，笛卡儿不应当从"我思故我在"之真假定"如此想的'我'是某物这一点是绝对必需的"。

[10]Williams（4），第95—100页。

[11]参见，例如，威廉斯在Williams（4）中的诘难。

[12]说我们没有什么有利于非还原论者的观点的证据过于独断。C·D·布罗德曾经明白地评述说，我们所拥有的证据，充其量支持下述结论：如果某个显然精神健全且严肃认真的人宣称具有相同种类的更好证据，那么，他的主张不应当简单地被忽视。

[13]这些说法脱胎于威廉斯的"想象和自我"，重印于Williams（2）。

[14]见Williams（8）。

[15]会有某些种类的非辨别性的（non-distinctive）连续性，诸如如何走路和奔跑的连续记忆。心理准则不应当诉诸这些种类的心理连续性。

[16]如前所述，至于对这些论证的讨论，请参见Dummett，Peacocke（1）、 Forbes（2）和Sainsbury。另请参见C.Wright"论模糊谓词的连贯性"，《综合》，1975年。至于对该论证应用于人的一种更长的讨论，请特别参见Unger （1） （2），以及P·安格尔在此一非凡系列中的其他作品。我尚未有时间考虑皮科克、福布斯和圣斯伯里所提出的解决之道，诉诸真理的程度，是否满足昂格尔的论证。

[17]Williams（2），第63页。

[18]《泰晤士报》，伦敦，1982年11月12日"科学专栏"。我被告知说，对于更加令人印象深刻的一些结果的报道将很快出现在适当的科学杂志上。

[19]格丽泰·嘉宝（Greta Garbo，1905—1990），美籍瑞典人，美国米高梅电影制片公司著名女星，以冷艳闻名。在无声电影和有声电影时期都有上佳表演。但是在事业和声誉的顶峰时期急流勇退，1941年后淡出电影圈。因其"令人难忘的银幕表演"，于1954年获奥斯卡特别奖。——译者

[20]Evans（1），第208页；另请参见Salmon，第245—246页。

[21]参见Broome（1）。

第十二章

我们的同一性为何不重要

87.分裂的头脑

近来的一些医学病例提供了有利于还原论者的观点的惊人证据。人的大脑有一个下脑和分成左右两个半球的上脑，这两个半球由神经束连接在一起。在治疗一些患有严重癫痫症的人的时候，外科医生就切断这些神经。目的是通过把癫痫发作的病因局限在一个脑半球上而减轻癫痫发作的强度。这个目的达到了。但是这样的手术产生了另一个无意的结果。用一个外科医生的话来说，这个结果就是创造了"两个独立的意识半球"。[1]

这个后果被各种各样的心理测试所揭示。这些测试运用了两个事实。我们用我们的左半脑控制我们的右手，用我们的右半脑控制我们的左手。我们视野中右半域中的东西是由左半脑看到的，而左半域中的东西是由右半脑所看到的。当某个人的左右半脑一经被分割开后，心理学家们就能够向这个人的视野中的两个半域出示写下的两个不同问题，而且能够获得这个人用两只手写下的不同的答案。

这里是这类试验所提供的此类证据的一个简化版本。这些人中的一位被带到一个宽屏幕面前，屏幕的左半部是红色的而右半部是蓝色的。在每个半部上都有颜色略暗的"你能够看到多少种颜色？"这么几个字。这个人用双手分别写下"只有一种"。屏幕上的字现在变成"你能够看到的惟一的颜色是什么？"这个人用一只手写下"红色"，用另一只手写下"蓝色"。

如果这就是这个人如何反应的，那么看来没有理由怀疑他正在拥有这样的视觉——正如他所声称的，他的确既看到红色又看到蓝色。但是当他在看到红色时没有意识到看到蓝色，相反亦然。这就是为什么那位外科医生说的"两个独立的意识半球"的原因。在他的每一个意

识中心，这个人只能够看到一种单一的颜色。在一个中心他看到红色，在另一个中心他看到蓝色。

许多实际试验，尽管在细节上与我刚才表述的想象的试验有所不同，但是表现出相同的两个本质特征。在看到他的视野中的左半域中的东西的时候，这个人没有意识到他正在右半域看到什么，反之亦然。在使他看到视野中的左半域而且意识到他在用左手干什么的那个意识中心当中，这个人没有意识到他正在用他的右手干什么，反之亦然。

实际事例中的一个复杂之处在于，对绝大多数人而言，至少在手术后的最初几个星期中，说话完全是由那个右撇子的脑半球控制的。结果，"如果'帽子'一词在左脑闪烁，如果这个人被告知挑出所看到的东西的话，左手将从一组被藏匿的对象中找出一顶帽子。同时他将口头上坚持说他什么也没有看到。"[2]另一个复杂之处在于，在经过一定时间之后，每个脑半球有时都能够控制双手。内格尔援引的一个此类例子如下：

> 一个烟斗在那个病人看不到的情况下被放到他的左手中，然后被要求用他的左手写下手中正握着的是什么。左手非常吃力和沉重地写下了字母P和I。然后突然间，写字速度加快起来而且轻松起来，字母I被转化成字母E，完成的字是pencil（铅笔）。显然左半脑基于最初显示的两个字母做出了猜测，而且进行了干预……但是随后右半脑重新接管了对那只手的控制，重重地划掉了encil这几个字母，而且画了一个烟斗草图。[3]

这样的冲突可能以更加危险的形式出现。这些病人中的一位抱怨说，有时候当他用右手拥抱自己的妻子的时候，他的左手却将她推开。

人们对在内格尔的例子中所触及的实际事例中的另一个复杂情况有了不少了解。左半脑典型地支持或者"具有"一个成年人的语言和数学能力，而右半脑在这方面"具有"的能力只有一个小孩子的水平。右半脑尽管在这些方面欠发达，但是具有其他方面的更大能力，诸如包

含在图案辨识方面或者音乐方面的那些能力。人们认定，在三四岁以后，两个半脑实行着一种"劳动分工"，各自发展出某些能力。右半脑较差的语言能力并不是固有的或者永久的。那些左半脑发生中风的病人语言能力经常退化到小孩水平，但是以其尚存的右半脑很多人能够重新习得成人的语言。人们还认为，少数人的两个半脑在这些能力方面没有任何差异。

假设我是这些少数人中的一位，有两个十分相似的半脑。而且假设我已经装备了能够切断我的两个半脑之间联系的某种装置。由于这个装置是与我的眉毛连在一起的，它处于我的控制之中。通过皱起一边的眉毛我能够分裂我的头脑（mind）。在我的分裂了的头脑的每一半，那么我能够通过松开我的一道眉毛来重新联合我的头脑。

这个能力会有许多用途。考虑一下

我的物理考试。我在参加一个考试，而且只剩下15分钟供我回答最后的问题。我想到解决这个问题的两个方法。我不能肯定哪一个更可能成功。我因而决定把我的头脑分裂10分钟，用我的头脑的每一半来尝试两种计算方法中的一种，然后再把我的头脑重新联合起来，写下最佳的修正结果。我将经历到什么？

当我切断我的两个半脑的联系的时候，我的意识流分裂了。但是这种分裂不是我体验到的某物。我的意识流中的每一个好像与我直到分裂那一刻的整个意识流直接相连。每个意识流中的惟一变化是我的半个视野的消失，以及对我的一个手的感觉和控制的消失。

考虑一下在我的"右撇子"意识流中的体验。我记得决定用我的右手做那个较长的计算。这是我现在开始的。在处理这个计算时，从我左手的运动我能够看出我还在处理另一个计算。但是我没意识到正在处理另一个计算。我或许在我的右撇子意识流中揣测在我的左撇子意识流中我是如何进行的。我会看和辨识。这就会像尝试着看清我的邻桌做的如何好一样。在我的右撇子意识流中我会同样既没意识到我的邻桌在想什么又没意识到我在我的左撇子的意识流中现在想的是什么。同样的评述适用于在我的左撇子意识流中的经验。

我的计算工作现在结束了。我即将重新联合我的头脑。在每一个意识流中我应该期待什么？简单地说，我将突然感到好像记得刚才在进行两个计算，当时在处理一个计算时没有意识到另一个。我提出，这我们能够想象。而且，如果我的头脑在此之前已经被分裂了，我的明显记忆会是正确的。

在表述这个事例时，我假定了有两个独立的思想和感觉系列。如果我的两只手可视地写出两个计算结果，而且我后来还主张记得两个

对应的思想系列，那么这就是我们应该假定的。假定两种计算中有一个或者两者都是无意识下进行的将会是最不可取的。

也许会有人诘难说，我的表述忽略了“意识的必然统一性”。但是我并未忽略这样的所谓的统一性。我已经否定了它。凡事实都必定是可能的。而左右半脑被隔断的那些人具有两个独立的意识流是一个事实——他们具有两套思想和经验，在具有其中的一个意识流时这两者都没有意识到具有另一个意识流。这两个意识流中的每一个都分别展示出意识的统一性。这可能是一个令人惊讶的事实。但是我们能够理解这个事实。我们能够渐渐相信一个个人的精神历程不需要像一个只有一股水道的沟渠，而是会像一条河流，有时具有分离的水流。我提出，我们还能够想象分裂和联合我们的头脑会是什么样子。我对我在物理考试中的经验的表述看来既是前后一贯的，又表述了我们能够想象的事。

接着或许会有这样的主张，认为在我的想象的事例中，我并非具有一个分裂的头脑。而是具有两个头脑。这个诘难并未真正提出问题。这些是描述同一个结果的两种方式。

一个类似的诘难主张说，在这些想象的和实际的事例中，结果不是要么具有一个分裂的头脑要么具有两个头脑的一个单一的人。结果是两个不同的人，分享控制一个躯体的绝大部分，但是每一个都单独控制一只手。此处我仍然认为这个诘难并未真正提出一个问题。这些再次是描述同一个结果的不同方式。如果我们是还原论者的话，这正是我们所相信的。

如果我们尚不是还原论者，正如我将假定的，我们认为有关这样的一些事例是否涉及到比一个单一的个人更丰富的东西的问题真正是一个问题。大概我们在实际的事例中能够相信这一点，在那里分裂是永久性的。但是当我们考虑想象出的我的物理考试的事例的时候，这个信念却是难以接受的。在该事例中，两个意识流只存在了10分钟。而且随后我好像记得做了两个计算，在那10分钟期间我的两只手能够

被看到写出了这些计算。鉴于这个非统一性其实是短暂的和不过量的，主张在这个事例中涉及到比一个单个的人更丰富的东西是不可取的。我们将要假设在这10分钟期间我停止存在，而且有两个新的人出现，每一个随后都做出了一个计算。照这个解释，整个一幕涉及到三个人，其中的两个人的寿命只有10分钟。此外，这两个人中的每一个都错误地认为他是我，而且具有与我的过去准确相符的明显记忆。在这10分钟之后，我具有对这两个人中每一个人的短暂生命的明显的准确记忆，但我错误地认为我自己曾经具有这两个人具有的所有思想和感觉这一点除外。对于此处我犯了错，而且这一幕的确涉及到三个不同的人，这是令人难以置信的。

同样令人难以置信的是，它涉及到两个不同的人，由我来做一个计算，由某个别人做另一个计算。我承认，当我最初分裂我的头脑的时候，我或许在做两个计算之一时认为另一个计算一定是由某个别人在做。但是在做另一个计算时我或许具有同样的信念。及至我的头脑被重新联合起来，随后我会好像记得相信在做一个计算的时候另一个计算在被另一个人做着。当这两个信念我好像都记得的时候，我会没有理由认为一个是真的，另一个是虚假的。而且在几次分裂和重新联合之后，我会不再具有这样的信念。在我的两个意识流中的每一个当中，我会相信我现在在我的另一个意识流中具有思想和感觉，对这些思想和感觉在这个意识流中我是意识不到的。

88.用什么解释意识的统一性?

假设因为我们还不是还原论者，我们认为对"谁有意识流？"这个问题一定有一个真正的答案。而且假设鉴于刚刚给出的理由，我们认为这个事例只涉及到一个单一的个人：我。我们认为在长达10分钟的时间内我具有一个分裂的头脑。

接着谨记心理的统一性由所有权来解释这一观点。照此观点，通过把不同的经验归属于一个人或者"经验主体"，我们应该解释这个人

的意识在任何时候的统一性。使这些不同经验统一起来的东西是他们正被同一个人所拥有这一点。这个观点既被那些认为个人是独立存在的实体的人们所坚持，又被拒斥这个信念的人中的某些人所坚持。而且这个观点还适用于每一个生命的统一性。

当我们考虑我想象的物理考试的时候，我们能够继续接受这个观点吗？我们认为，当我的头脑被分裂的时候，我有两个独立的经验系列，在具有其中一个的时候我意识不到具有另一个。在我的一个意识流中的任何时候，我都具有数个不同的思想和感觉。我可能意识到思考计算的某个部分，意识到感觉书写者的一只手的抽筋，意识到听见我的邻座老式钢笔的沙沙声。是什么使这些不同的经验统一起来？

按照上述观点，答案是这些是我在这个时刻正在拥有的经验。这个答案是不正确的。在这个时刻我并不就具有这些经验。在我的另一个意识流中还具有数个其他的经验。我们需要解释我的两个意识流各自的内部的意识的统一性，或者我的分裂的头脑的每个半球内的意识的统一性。我们不能够通过主张所有的这些经验都是这个时刻为我所拥有来解释这两个统一性。这使两个统一性成为一个统一性。它忽略了这样一个事实，即在具有每一组经验的时候我没有意识到具有另一组经验。

假设我们继续认为统一性应该通过把不同的经验归属于一个单一的主体来解释。那么我们必定认为这个事例涉及到至少两个不同的经验主体。统一起我的左撇子意识流中的那些经验的东西是它们都为一个经验主体所拥有。统一起我的右撇子意识流中的经验的东西是它们都为另一个经验主体所拥有。我们现在必须放弃"经验主体"是个人这一主张。按照我们的观点，我是一个经验主体。而当我的头脑被分裂的时候，有两个不同的经验主体。它们不是同一个经验主体，所以它们不能两者都是我。既然我不大可能是两者中的一个，有鉴于我的两个意识流的相似性，我们大概应该断定我不是这两个经验主体中的任何一个。整个情节从而涉及到三个这样的实体。这些实体中的两个不

能当成那类我们都熟悉的实体——个人。我是涉及到的惟一的个人，而且这些经验主体中有两个不是我。即使我们假定我是这两个经验主体中的一个，另一个也不能是我，因此也不是一个个人。

我们现在可能是怀疑论的。当"经验主体"是个人的时候，主张统一起一组经验的是它们都为一个单一的个人所拥有这一点好像是令人信服的。即便我们不得不认为有一些不是个人的经验主体，我们仍然可能怀疑它们是否真是这样的东西。当然，在动物世界中有许多不是个人的经验主体。我的猫就是其中的一例。但是其他动物则与这个想象的事例无关。照上述观点，我们不得不认为一个个人的生命会涉及到不是个人的一些经验主体。

重新考虑一下在我的右擨子意识流中的经验。在这个意识流中的某个时刻，我意识到思考计算的某个部分，意识到感觉书写者的一只手在抽筋，意识到听见我的邻座老式钢笔的沙沙声。我们通过主张这些经验都是被那同一个经验主体——不是我的一个实体——所拥有，来解释这些经验的统一性吗？这个解释好像并不可取。如果这个经验主体不是一个个人，那么它是哪一类东西？它不能够被当成是一个笛卡儿式的自我，如果我被当成是一个笛卡儿式的自我的话。这个经验主体不能够被当成是一个这样的自我，因为它不是我，而且这个事例只涉及到一个个人。这个经验主体能够是一个笛卡儿式的亚自我，只是作为个人的一部分的一个持久的纯精神实体吗？我们可能决定我们没有充分的根据认为有这样的事物的存在。

下面我转到上述的另一个观点。一些人认为统一性是通过所有权来解释的，即使他们否定我们是独立存在的实体。这些人认为，在任何时刻统一起个人的经验的东西是基于这样的事实，即这些经验正被这个人所拥有。正如我们已经看到的，在这个想象出的事例中这个信念是虚假的。当我在我的右擨子意识流中具有一组经验的时候，我还在我的左擨子意识流中拥有另一组经验。我们不能通过主张这些经验

为我现在所拥有来解释任何一组经验的统一性，因为这会把这两组经验合并起来。

一个还原论者现在可能介入进来。按照他的观点，统一起我右撇子意识流中的经验的东西是：任何时候都有一个单一的对这些不同的经验的意识状态，有一个具有某些思想、感到书写者痉挛和听到钢笔沙沙声的意识状态。同时，在我的左撇子意识流中有另一个各种各样经验的意识状态。我的头脑是分裂的，因为没有关于这两个经验组的一个单一的意识状态。

或许有人提出诘难说，这些主张并未解释每一个意识流中经验的统一性，而只是重新描述了每一个意识流中的经验统一性。在某种意义上说，确实如此。这个统一性并不需要深层的解释。它只是这样一个事实，即数个经验能是共意识的，或者是一个单一的意识状态的对象。把这个事实与对最近时刻的经验的短期记忆——对所谓的"似是而非的在场"的短期记忆——这一事实加以比较，可能是有益的。就像对刚刚具有的数个经验，诸如听到钟响三声，能够有一个单一记忆一样，对听到钟响第四声和看到乌鸦掠过钟楼这两者能够有一个单一的意识状态。还原论者们主张，在某个单一时刻意识的统一性中再没有涉及到什么更多的东西了。既然对数个经验能够有一个意识状态，我们就不需要通过把这些经验归属于同一个个人或者经验主体来解释这个统一性。

值得重申还原论者的观点的其他一些部分。我主张：

> 由于我们把思想归属于思想者，所以思想者的存在为真。但是思想者不是独立存在的实体。一个思想者的存在只涉及他的大脑和躯体的存在，以及他的行动的实施、他的思想的思考、其他某些物理和精神事件的发生等的存在。因此我们能以非人称的术语来重新描述任何个人的生命。在解释这个生命的统一性的时候，我们不需要主张它是某个特定的个人的生命。我们能描述在不同的时刻所想、所感、所看和所做的事情，以及这各种各样的事件是如何互联在一起的。只有在描述许多思想、愿望和记忆等等的内容的时候，个人在此处才会被提及。个人不需要被当作这些思想中任何一个思想的思想者。

这些主张得到其中我分裂我的头脑的那个事例的支持。不同经验的统一性不需要通过把所有的这些经验归属于我来解释这一点此处不仅为真，而且我在每一个意识流中的经验也是不能以这种方式被解释的。只有两种选择。我们可能把每一个意识流中的经验归属于一个不是我、因而不是一个个人的经验主体。或者，如果我们怀疑这样的一些实体的存在，我们可以接受还原论者的解释。至少在这个事例中，现在看起来这可能是最好的解释。

这是我想象出的事例是否可能变得重要起来的那些关键处之一。如果我们会暂时地分裂我们的头脑的话，这给心理的统一性由所有权来解释这一观点投下了怀疑的阴影。正如我所论证过的，如果我们不是还原论者的话，我们应该把我想象的事例当作只涉及一个单一的个人来对待。那么主张意识的统一性应该通过把不同的经验归属于一个单一的主体——个人——来解释就变得不可能了。只有通过认为那些经验主体不是个人，我们才能够坚持这个观点。（人之外的）其他动物在此是无关的。与我们的信念有关的是在个人的生命中涉及的是什么。如果我们不得不承认在这些生命中会有两类经验主体——那些是个人的经验主体和另一些不是个人的经验主体的话，我们的观点的令人信服之处将要大打折扣。如果我们主张，因为个人是不可分的，所以我想象的事例永远不会发生的话，这会有助于我们的观点。

我的事例是想象的。但是该事例的本质特征，意识分裂成独立的意识流则已经发生了许多次。这削弱了刚才给出的回应。我所想象的事例或许变得可能起来，而且最多只是技术上不可能。而且在这个事例中每一个意识流中意识的统一性不能通过把我的经验归属于我而得到解释。因为这个解释站不住脚，这个事例反驳了心理统一性能够通过把不同的经验归属于一个单一的个人来解释的观点。

按照这个观点的最为著名的版本，我们是笛卡儿式的自我。我对利希滕贝格对笛卡儿式的观点的诘难进行过辩护。但是那个辩护只是表明我们不能从我们经验的本质推导出我们是这样的实体。我后来主

张并没有有利于这个观点的证据，而且相当的证据不利于它。既然那些分裂头脑的实际事例支持刚刚给出的论证，那么它们就是些针对这个观点的进一步的证据。

笛卡儿的观点可以与牛顿的绝对时空观相比较。牛顿认为，任何物理事件只是通过它与时空这两个独立实在的关系而具有一个特定的位置。我们现在认为，一个物理事件借助它与其他发生的物理事件的各种各样的关系而具有一个特定的时空位置。按照笛卡儿式的观点，一个特定的精神事件只是借助它被归属于一个特定的自我而在一个特定的生命中发生。我们能够否定"精神空间"是由这样的持久诸自我的存在所给予的。我们能够主张，一个独特的精神事件借助它与（通过互联而构成这个生命的）许多其他精神和物理事件的那些关系而在某个生命之内发生。[4]

有时人们还给出相信这样的诸自我的另一个根据。针对任何完全客观的实在表述——任何不是从某个"观点"所作的表述，人们能够认为它忽略了某些真相。一个相关例子是，我是我，或者我是德里克·帕菲特。我是这个特定的个人。这些主观的真相看来可能蕴含着我们是独立存在的经验主体。

这样的真相能够被一个还原论者所陈述。"主观的"一词容易令人误入歧途。被称作主观真相的东西不需要涉及任何经验主体。一个特定的思想可能是自指的。它可能是这样一个思想，即这个特定的思想即便与其他被思的思想确切相似仍然是这个特定的思想——或者是对这个思想的特定思考。这个思想是一个非人称的，但却是主观的真相。

一些人会诘难说，所有的其他指标性的概念——诸如"这里"、"现在"和"这个"——必须以一种使用概念"我"的方式来解释。实际并非如此。其他所有的指标性概念，包括"我"在内，能够以一种使用"这个"的自指用法的方式来解释。而且这个自指用法并不涉及到一个自我观念，或者一个经验主体。正是在这个句子中用"这个"来指这个句子。

以“这个”的这个用法，我们不需要相信经验主体的独立存在就能够表达“主观的”真相。[5]

89.我分裂时发生了什么

现在我将描述有关头脑分裂的另一个自然延伸的实际事例。假设我是一对同卵孪生兄弟中的一个，我的躯体和我的孪生兄弟的大脑都受了致命创伤。由于神经外科领域的进步，这些创伤将不会不可避免地导致我们两个都死亡。我们之间有一个健康的大脑和一个健康的躯体。外科医生们可以把它们放到一起。

即使用现有的技术这一点也是能够做到的。正如我的大脑能够被摘取出来并且通过与一个人工心肺机相联而保持活体状态一样，我的大脑也能够通过与我的孪生兄弟躯体内的心和肺的相联而保持成活。现今的缺点是我的大脑的神经还不能与我的孪生兄弟躯体内的神经相连接。我的大脑如果移植到他的躯体内的话会成活，但是作为结果的人会是瘫痪的。

纵然他是瘫痪的，作为结果的人会能够与他人交流。一个粗糙的方法会是把某个装置固定到那个控制这个人的右拇指的神经上，使他能够用莫尔斯电码传送信息。另一个装置固定到某个感觉神经上，会使他能够接收信息。很多人会对幸存持欢迎态度，即使是完全瘫痪的一种劫后余生，只要他们仍然能与其他人保持交流。一个现成的例子是一个伟大的科学家，他的生活的主要目标是继续思考某些抽象的难题。

不过，让我们假设外科医生能够把我的大脑与我的孪生兄弟躯体内的神经相连接。作为结果的人就不会瘫痪，就会是完全健康的。这个人会是谁?

这不是一个困难的问题。看起来物理的和心理的准则在此处可能存在分歧。尽管作为结果的人将在心理上与我连续，但是他将不具备

我的整个躯体。但是，正如我已经主张的，物理准则不应该要求我的整个躯体的连续存在。

如果我的大脑全部连续存在并且是一个在心理上与我连续的某个活生生的人的大脑，那么我将连续存在。无论我（原来）的躯体的其他部分发生了什么变化，这都为真。当我被赋予某个别人的心脏的时候，我是幸存下来的接受者，而不是死去的捐献者。当我的大脑被移植到某个别人的躯体中的时候，可能看起来这里我是死亡了的捐献者。但是我仍然是真实的接受者，而且是幸存者。接受一个新头脑、一个新躯体正是接受新心脏、新肺脏和新手臂等等的极限事例。[6]

我的新躯体像什么将当然是重要的。如果我的新躯体根本不像我的旧躯体，这会影响到我能干什么，而且或许会间接导致我性格上的变化。但是没有理由假设被移植到一个非常不同的躯体这一点会瓦解我的心理连续性。

有人已经提出诘难说，“拥有某些种类的性格特性要求拥有一种适当的躯体”。昆顿回答了这个诘难。他写到一个不太可能的事例：

> 一个6岁的小女孩展现出温斯顿·丘吉尔的性格会是件怪异的事，达到坚忍不拔这一点的确是怪异的，但不是完全不可想象的。毫无疑问，小女孩所展现的顽强忍耐力、对世界和历史的宏观把握等等，起初会让人感到在一个这么小的孩子身上会有些味道不对和自命不凡。如果她把这一点保持下去，这样的印象就会渐渐淡去。[7]

更加重要的是，正如昆顿所论证的，这个诘难只表明重要的或许是我的大脑是否落户在某种躯体中，但是它并不表明重要的是我的大脑是否落户在任何特定的躯体中。而且在我想象出的事例中，我的大脑将落户在这样一个躯体中，这个躯体尽管在量上并不与我的旧躯体同一，却非常相似——因为这是我的孪生兄弟的躯体。

照心理准则的所有版本来说，作为结果的人都会是我。而且绝大多数信奉物理准则的人们会被说服，承认在这个事例中此为真。正如我已经主张的，物理准则应该只要求我的足量作为一个活生生的人的

大脑的那部分大脑的连续存在。前提是没有任何别人具有这样的大脑。这会使那个手术后苏醒过来的人是我。而且如果我的孪生兄弟的躯体就像我的躯体一样，我甚至可能不会注意到我有了一个新躯体。

真的，事实上一个半脑就足够了。当脑中风或者意外伤害使许多人的两个半脑中的一个失去功能时，他们幸存了下来。以其尚存的半脑，这样的人可能需要重新习得某些东西，诸如成人的言语和如何控制双手等等。但这是可能的。在我所举的例子中，我假定就像在许多实际生活中的人的真实情况一样，我的两个半脑都各自完全具有一系列的能力。这样我会以其中的任何一个半脑幸存下来，而无需重新习得。

现在我将把这最后两个主张结合起来。如果我的大脑被成功地移植到我的孪生兄弟体内的话，我也会幸存下来。而且在我的大脑的一半业已被毁的情况下，我会只以我的大脑的另一半而幸存下来。鉴于这两个事实，显然，如果我的大脑的一半被成功地移植到我的孪生兄弟的体内，而另一半被毁，我亦会幸存下来。

如果另一半大脑不被毁掉又当如何？这正是威金斯所表述的一个事例，其中一个人像变形虫一样分裂了。[8]为简化该事例起见，我假定我是同卵三胞胎中的一个。考虑一下

> 我的分裂。我的躯体受了致命伤，就像我的两个兄弟的大脑受了致命伤一样。我的大脑被分割成两半，分别成功地移植到我的两个兄弟的体内。作为结果的两个人中的每一个都认为他是我，好像记得过去我的生活，具有我的性格，而且其他每一个方面都在心理上与我连续。而且他具有一个与我的躯体非常相似的躯体。

这个事例很可能仍是不可能的。尽管有主张说，在某些人中两个半脑可能具有完全一样的能力，但是这个主张或许是虚假的。我在这里假定这个主张在应用到我身上的时候为真。我还假定一个被移植的半个大脑与其新躯体中的神经相连接会是可能的。而且我假定我们能够分裂，不仅是指脑上部的两个半球，而且脑下部也能够分裂。如果

神经生理学取得足够进步的话，我的前两个假定可能会成为现实。但是以一种不损坏其功能的方法来分割下脑，看来好像是永远不大可能的。

如果由于这个理由，这个想象的完全分裂的事例将总是不可能的，那么重要吗？鉴于我的讨论的诸目标，这并不重要。这种不可能纯是技术上的。或许被当作根本上不可能的一个该事例的特征——一个人的意识分裂为两个独立的意识流——是一个实际上已经发生了的特征。即使本不可能的话，它本来也是重要的。因为这可能已经支持某个有关我们真正地是什么的主张了。它可能已经支持我们是不可分的笛卡儿式的自我这一主张了。因而一个人的意识分裂事实上是可能的这一点是重要的。在我们真正是什么这样一个特定观点与分割下脑、成功移植下脑的两半之不可能性之间，好像没有什么相似的关系。于是这种不可能性并不为拒绝考虑那个想象出的事例提供根据，在那个事例中我们假设这是能够做到的。而且考虑这个事例可能既有助于决定我们相信我们是什么，又有助于决定事实上我们是什么。正如爱因斯坦的例子所表明的，考虑一些不可能的思想-实验会是有益的。

事先陈述这个事例表明什么，可能有所助益。它进一步论证了有关我们是独立存在的实体的观点。但是有待得出的主要结论是个人同一性无关紧要。

认为我们的同一性重要是自然的。重新考虑一下那个支线事例，其中我已经与我在火星上的复制品谈了话，而且将要死去。假设我们认为我和我的复制品是不同的人。那么假定我的前景几乎像通常的死亡一样糟就是件自然的事。几天之内，将不会有一个是我的人在生活。假定这是重要的是件自然的事。在讨论我的分裂的时候，我将从作出这一假定开始。

在这个事例中，我的大脑的两个半脑将分别成功地移植到我两个兄弟十分相似的躯体内。两个作为结果的人将在心理上完全与现在这

样的我连续。我身上发生什么?

只有四种可能性：（1）我没有幸存下来；（2）我作为两个人中的一个幸存下来；（3）我作为两个人中的另一个幸存下来；（4）我作为两者都幸存下来。

对（1）的诘难是这样的。如果我的大脑被成功地移植的话，我会幸存下来。而且事实上有些大脑损坏了一半的人幸存了下来。鉴于这些事实，如果我的一半大脑被成功地移植而另一半大脑被毁掉，我会幸存这一点好像很明显。这样，如果在我的另一半大脑也被成功移植的情况下，我又怎么不能幸存下来呢?

考虑一下接下来的两个可能性。成功大概就是最大程度的成就。也许我是两个作为结果的人中的一个。此处的诘难是：在这个事例中我的大脑的两个半脑是确切相似的，而且一开始两个作为结果的人也是如此。鉴于这些事实，我怎么能只是作为两个人中的一个而幸存下来呢？是什么能够使我成为其中的一个而不是另一个?

这三种可能性不能够因其不连贯而被拒绝考虑。我们能够理解它们。但是当我们假定同一性重要的时候，（1）是不可取的。我的分裂不会像死亡一样糟。（2）和（3）也是不可取的。剩下第四种可能性：我作为两个作为结果的人双双幸存下来。

这种可能性也许可以用数种方式加以描述。我可能首先主张："我们所谓的'作为结果的两个人'不是两个人。他们是一个人。我的确经历这个手术后幸存下来。其后果是赋予我两个躯体和一个分裂的头脑。"

这个主张不能直接加以摒弃。正如我论证过的，我们应该承认一个人会有一个分裂的头脑这一点是可能的。如果这是可能的，那么我的分裂头脑的每一半可能可以控制其自己的躯体。尽管对该事例的这个描述不能被当作不可想象的东西予以拒斥，但是它涉及到个人概念上的一个巨大的变化。在我想象出的那个物理考试事例中，我主张这个事例只涉及到一个个人。该事例的两个特征使之具有可取性。那个分裂的头脑很快得到重新统一，而且只有一个躯体。如果一个头脑被

永久地分裂，而且其两半各自以不同的方式发展，主张该事例只涉及到一个人就会变得不合情理。（记住那个病人的实例，他抱怨拥抱自己的妻子时自己的左手把她推开。）

也涉及两个躯体的那个完全分裂的事例，好像远远超出了那个边际。在我做完手术之后，两个"产品"每一个都有一个人的所有特征。他们可能会生活在地球的两端。假设他们的记忆力不强，而且他们的外貌发生了不同的变化。在许多年之后，他们或许重新相遇，而且甚至没能相互认出对方。我们可能不得不就这一对浑然不觉地在打网球的人主张说："你在那里所见到的是某个单一的个人，在同他自己打网球。在他的头脑的每一半中，他都错误地认为他正在同某个别人打网球。"如果我们尚不是还原论者，我们认为对于这两个打网球的人是否是一个单一的个人有一个真实的答案。有鉴于我们的"个人"一词的含义，答案一定是否定的。我相信是个陌生人的那个站在球网后面的人事实上却是我自己的另一部分，这一点不能为真。

假设我们承认那两个"产品"，就像他们看上去的那样，是两个不同的人。我们还会主张我作为两者幸存下来吗？我们能够另辟蹊径作如此主张。我会说："我作为两个不同的人在手术后幸存下来。他们可以是两个不同的人，但是仍然是我，就像教皇的三个皇冠一起组成一个皇冠一样。"[9]

这个主张也是前后一致的。但是它再次大大曲解了个人的概念。我们欣然同意，教皇的三个皇冠被放到一起时形成第四顶皇冠。但是难以把两个在一起的人看成是一个第三人。假设作为结果的人进行决斗。那么，这是一边一个人，一边两个人的决斗吗？而且假设发射的一粒子弹打死了人。这是一个谋杀与一个自杀的两种行为吗？多少人活下来了？一个还是两个？这个复合的第三人没有独立的精神生活。令人难以置信，实际上会有这样一个第三人。与其说作为结果的人共同构成我——以至于二重唱变成三重唱——不如把他们当作一对而且以一种更简单的方式表述他们与我的关系来得更好些。

人们可能会提出一些别的主张。可能会提出，两个作为结果的人现在是不同的人，但是在我的分裂之前他们曾是同一个人。在我的分裂之前，他们曾是我。这个解释是模棱两可的。该主张会说在我的分裂之前他们一起是我。照此说法，甚至在我的分裂之前就曾有三个不同的人。这比我刚刚拒斥的那个主张更加缺乏令人信服之处。（或许有人认为我误解了这一解释。该主张可能是说作为结果的人在我的分裂之前作为独立的人并不曾存在。但是，如果他们那时并不存在，那么他们曾一起是我就不能为真。）

人们可能转而提出在我的分裂之前每一个作为结果的人曾是我。在我的分裂之后，两个都不是我，因为我现在并不存在。但是，如果每一个这样的人曾是我，那无论我身上发生了什么，都一定已经发生在每一个这样的人身上。如果我在我的分裂之后不曾幸存下来，那么这些人没有一个幸存下来，既然现有两个作为结果的人，该事例涉及到五个人。这个结论是荒谬的。我们能够否定蕴含着这个结论的假定吗？我们能够主张尽管每一个作为结果的人曾是我，但是曾发生在我身上的并未曾发生在这些人身上吗？假定我还没有分裂。照此来说，每一个作为结果的人现是我现在为真。如果发生在我身上的并不发生在X身上，X不能是我。

否定最后这个主张有一些牵强的方式。这些方式诉诸有关"时态性的同一性"。姑且称一个作为结果的人为老左。我或许问，"老左和德里克·帕菲特是同一个人的名字吗？"对那些相信时态性的同一性的人们而言，这不是一个适当的问题。正如这表明的，有关时态性的同一性的主张与我们现在所思考的方式极为不同。在此我只想陈述我认为别人已经表明的一点：这些主张并不能解决我们的难题。

大卫·刘易斯提出了一个不同的建议。照他的观点，甚至在我的分裂之前就有两个人分享我的躯体。在细节上，这个建议既雅致又机智。我在此将不重述为什么，就像我在别处已经主张的那样，这个建议并不解决我们的难题。[10]

我们已经讨论了有关我分裂时发生什么的数个异乎寻常的观点。照这些观点，该事例牵涉到一个单一的人的独唱，一对二重唱，一组三重唱（其中两人组成第三人）和一组五重唱。我们毫无疑问会猜想出那个缺席的四重唱。但是考虑这些观点之外的更多的观点会显得冗长乏味。所有的观点都牵涉到对个人概念的巨大曲解。因而我们应该拒斥所提出的第四个可能性：即在某种意义上我是以两个作为结果的人而双双幸存下来的。

还有其他三种可能性：我将是这些人中的一个、另一个，或者不是两者中的任何一个。这三个主张曾看起来并不可取。下面请注意，就像以前，就算我们实际能操作这个手术，我们也不能发现什么。假设，例如，我的确作为一个作为结果的人而幸存下来。我会认为我已经幸存下来。但是我会知道另一个作为结果的人会虚妄地认为他是我，而且是他幸存下来。既然我会知道这一点，就会不信赖我自己的信念。我会是具有虚妄信念的那个作为结果的人。而且，既然我们两者都主张是我，他人就会没有理由认为我们中的一个而不是另一个。从而，就算我们实施了这个手术，我们也不会有所收获。

无论在我身上发生了什么，我们也不能发现发生了什么。这为我们的问题提出了一个更加极端的答案。它提示着还原论者的观点为真。这里大概没有不同的可能性，其中的每一个或许是所发生的，尽管我们永远不会知道实际上发生的是哪一个。大概，当我知道每一个作为结果的人会有我的大脑的一半，以及会与我在心理上连续的时候，我们也就知道了一切。例如，当我们提出作为结果的人中的一个或许是我的时候，我们在假设什么？什么会使这成为真正的答案？

我认为不能有不同的可能性，其中的每一个或许都是真相，除非我们是像笛卡儿式的自我那样的独立存在的实体。如果我所示的真是一个特定的自我，这就解释了作为结果的人中的一个是我这一点如何会为真。正是在这个人的大脑和躯体中这个特定的自我重获意识，这会是真的。

如果我们相信笛卡儿式的自我，我们或许会想起伯里丹的驴（Buridan's ass）这一典故，这头驴面对同样富有营养的两捆干草而被饿死。这头驴在吃这两捆干草中另一捆之前没有理由吃这干草中的这一捆。作为一个过分理性的牲畜，它拒绝作出没有理由的选择。在我的例子中，为什么我所示的那个特定的自我应该作为两个作为结果的人中的一个而苏醒过来会没有理由。但是这或许就是随机地发生的，就像基本粒子一样。

对笛卡儿式的自我的信奉者们而言更加困难的问题是，我到底是否会幸存。既然每一个作为结果的人会在心理上与我连续，就不会有证据支持对于这个问题的任何一个答案。即使我是一个笛卡儿式的自我，这个论证也具有力量。

就像以前，一个笛卡儿的信奉者可能诘难说，我误述了所发生的事。他可能主张，如果我们进行这个手术，作为结果的人两者都会在心理上与我相连续事实上不会为真。这些人中这一个或者那一个在心理上与我相连续或许为真。在这些事例中的任何一个中，这个人会是我。也许两者中没有一个在心理上与我相连续反而为真。在这种情况下，我不会幸存下来。在这三种情况中的任何一种情况下，我都会获得真相。

这是否是一个好诘难取决于我们的心理特征和我们的大脑之间的关系是什么。正如我已经说的，我们具有有关心理连续性的承载者并不是不可分的决定性证据。在左右半脑业已被割裂的那些病人实例中，产生了两个系列的思想和感觉。这两个意识流在心理上都与原先的意识流相连续。心理的连续性于是已经在数个实例中呈现出一个分裂的形式。这个事实驳斥了刚才的诘难。它证明我的这一个主张是正当的，即在我的分裂这个想象的事例中，两个作为结果的人都会在心理上与我相连续。既然如此，能够被提出的笛卡儿式的观点也只能是一种更加令人怀疑的版本，这个版本并不把自我与任何可观察的或者可内省的事实相连接。即使我是这样的一个自我，我也决不会知道我

是否已经幸存下来。对笛卡儿的信奉者们而言，这个事例是一个不可解的难题。

假设，鉴于早先给出的理由，我们拒斥有关我们每一个人其实都是一个笛卡儿式的自我的主张。而且我们拒斥有关个人是与其大脑、躯体和各种各样的精神、物理事件相分离的其他任何种类的独立存在的实体的主张。那么，我们应该如何回答有关我分裂时发生什么的问题？我区分过四种可能性。当我讨论每一种可能性的时候，好像都有一些针对该可能性就是会发生的情况这一主张的诘难。如果我们相信这四种可能性是一些不同的可能性，其中的任何一种都可能是所发生的情况，那么该事例对我们来说就也是一个难题。

依照还原论者的观点，这个难题消失了。照此观点，我已经讨论过的那些主张并不描述不同的可能性，其中的任何一个可能为真，而且其中的一个必定为真。这些主张只是对同一结果的不同描述。我们知道这个结果是什么。将有两个未来的人，其中的每一个将拥有我的一个兄弟的躯体，而且将在心理上与我相连续，因为他具有我的大脑的一半。知道了这一点，我们就知道了一切。我可能问，"但是我将是这两个人中的一个呢，还是另一个呢，抑或哪一个也不是？"但是我应该视之为一个空洞的问题。下面是一个类似的问题。在1881年，法国社会党分裂。发生了什么？法国社会党不再存在了吗？还是作为两个新的党的一个或者另一个而继续存在呢？鉴于某些进一步的细节，这会是一个空洞的问题。即使我们对这个问题没有答案，我们会知道究竟发生了什么。

我现在必须区分使一个问题可能空洞的两种方式。关于一些问题，我们应该主张它们既是空洞的又是没有答案的。我们会决定赋予这些问题答案。但是任何可能的答案都会是随意的这一点可能为真。如果是这样，给出此类问题以答案这一点会是无的放矢的，而且可能容易令人误入歧途的。联合谱系中那些主要事例中的那个"我将幸存

吗？”的问题或许就是如此。如果在远端事例中我不会幸存的话，在其他谱系的主要事例中上述这点也适用。

还有另一类其中问题可能空洞的事例。在这样的一个事例中，这个问题在某种意义上有一个答案。该问题是空洞的，因为它并不表述不同的可能性，其中的任何一个可能为真，而且其中的一个必定为真。该问题只是赋予我们同一个结果的不同表述。我们不经选择这些表述中的一个就会知道有关这个结果的整个真相。但是，如果我们的确决定赋予这个空洞的问题一个答案，那么这些表述中的一个就比另一些要好些。既然如此，我们能够主张这个描述是对这个问题的一个回答。而且我主张就我分裂的那个事例而言有一个最佳的描述。最佳的描述就是，作为结果的两个人中将没有哪一个是我。

既然这个事例并不涉及到不同的可能性，重要的问题不是“哪一个是最佳描述？”而是“什么应该对我来说是重要的？我应该如何看待分裂？我应该把它当作像死亡一样，还是当作像幸存一样？”当我们已经回答了这个问题的时候，我们能够决定我是否已经给出了最佳描述。

在讨论什么是重要的之前，我将兑现一个早先的承诺。对心理准则的一个诘难是，心理连续性预设了个人同一性。在记忆事例中，通过诉诸那个更加宽泛的准记忆概念，我回答了这个诘难。珍妮的类记忆记得具有某个别人的过去经验。我的分裂事例提供了另一个例子。既然至少两个作为结果的人中的一个将不是我，他能够类记忆地记得过着某个别人的生活。

我并未表明在表述心理连续性中所牵涉到的其他关系时我们需要预设个人同一性。既然我已经表述了我的分裂，这点就能够容易地得到表明。另一个直接的关系是某个意图和随后把这个意图付诸实践的行动之间的关系。我们可能试图只采取我们自己的行动，这一点可能在逻辑上为真。但是我们能够使用一个新的概念——类意图。一个个人可能会接近打算采取另一个人的行动。当这个关系成立的时候，它并不预设个人同一性。

那个分裂事例表明这涉及到什么。我能虚拟打算一个作为结果的人周游世界，同时另一个人呆在家里。我所虚拟打算的将不是由我来做，而是由两个作为结果的人来做。正常情况下，如果我打算某个别人应该做某事的时候，我只是通过形成这样的意向，而不能够使其做这事。但是如果我那时将要分裂，只是形成虚拟意向就会足够了。两个作为结果的人都会继承这些虚拟意向，而且，除非他们改变了他们所继承的想法，他们会把它们予以实施。因为他们或许改变他们的主意，所以我就不能肯定他们会做我所虚拟意向的。在我自己的生活内同理亦然。因为我可能改变我自己的主意，所以我不能肯定我将做我现在所打算去做的事。但是我具有通过形成坚定的意向来控制我的未来的某种能力。如果那时我将要分裂，我会刚好具有那么多的能力，通过形成虚拟意向来控制两个作为结果的人的未来。

相似的评述适用于所有其他的直接心理联系，诸如在性格连续性中所牵涉到的那些联系等等。所有的这样的联系处于我和每一个作为结果的人之间。既然至少这些人中有一个不能是我，这些联系中就没有任何联系预设个人同一性。

90.我分裂时重要的是什么

一些人会把分裂当作如同通常的死亡一样糟，或者几乎同通常的死亡一样糟。这种反应是非理性的。我们应该把分裂当作几乎与通常的幸存一样好。正如我已经论证的，这个手术的这两个“产品”会是两个不同的人。考虑一下我与这两个人中的每一个人之间的关系。这样的关系未包含在通常的幸存中包含的某个生死攸关的要素吗？显然不是没有。只要我与这些结果的人中的一个处于这样相同的关系中，我就会幸存下来。纵然一个人的大脑的一半被毁他也能够幸存，这是一个事实。而且经再三反思，如果我的整个大脑被成功地移植到我兄弟的体内，我会幸存这一点是明显的。从而，如果我的半个大脑被毁而另半个大脑被成功地移植到我兄弟的体内的话，我会幸存这一点也是

明显的。在我们现在所考虑的事例中，我与作为结果的人中的每一个人之间的关系，于是包含着对我作为那个人幸存下来所需要的一切。在这个事例中，不可能是我与作为结果的人中的每一个人之间的关系的性质导致它不是幸存关系。什么都不缺少。错误之处只能是在复制这一点上。

假设我接受此点，但是仍然把分裂当作几乎与死亡一样糟。现在我的反应是不可辩护的。我就像某个被告知一种药物能够延长寿命一倍的人，却把服用这种药物当作死亡。在分裂事例中的惟一区别在于，额外的那些年份是同时并存度过的。这是一个有趣的差异；但是这不能意味着没有年份去度过。我们或许说："你将失去你的同一性。但是有这么做的不同的方式。死亡是其中之一，分裂是另一种方式。把这些等量齐观，无异于混淆了2和0。双重幸存不同于通常的幸存。但是这并不使双重幸存成为死亡。双重幸存甚至更加不像死亡。"

双重幸存的难题在于它并不符合同一性的逻辑。就像还原论者们一样，我主张

> 关系R重要。R是有那类正确原因的心理联系性和／或者心理连续性。[11]

我还主张

> 在对于什么重要的说明中，那类正确原因可以是任何原因。

其他的还原论者们或许要求R有一个可靠的原因，或者要求有其正常的原因。我们暂缓处理这个分歧，只考虑R会有其正常原因的事例。在这些事例中，还原论者们都会接受下述这一主张。如果一个未来的人将与现在所是的我有关系R，而且没有不同的人与我有关系R，这个人就将是我。如果没有这样一个不同的人，这个未来的人将是我这个事实就只寓于我们之间有关系R这一事实中。就个人同一性而言只不过是关系R之持有。在几乎所有的实例中，R都呈现出一对一的形

式。它处于一个现在存在的个人与一个未来的个人之间。当关系R呈现出一对一的形式时，我们能够使用同一性的语言。我们能够主张这个未来的人将是这个现在的人。

在想象出的那个我的分裂的事例中，R呈现出的是一种分叉形式。但是个人同一性不能呈现一种分叉形式。我和两个作为结果的人不能是同一个人。既然我不能与两个作为结果的人相同一，称其中的一个人是我就会是武断的，我们充其量只能够通过说两者中没有哪一个是我来描述这个事例。

哪一种关系更重要？是个人同一性还是关系R？在通常的情况中，我们不需要决定其中的哪一个重要，因为这些关系刚好重合。在我的有关分裂的那个事例中，这些关系并不重合。因此我们必须决定两者中哪一个重要。

如果我们认为我们是独立存在的实体，我们就能令人信服地主张，同一性是重要的。照此观点，个人同一性是一个深层的进一步的事实。但是我们有充分的证据来拒斥这个观点。假如我们是些还原论者的话，我们就不能够令人信服地主张，就这两个关系而言，是同一性重要。照我们的观点，当关系R呈现为一种非分支形式的时候，个人同一性的事实就寓于关系R的持有之中。如果个人同一性就寓于这另外一个关系中，这另外一个关系必定是重要的。

人们可能诘难说："你主张对同一性而言只不过是关系R这一点是错误的。正如你已经说的，个人同一性具有一个额外的特征，并未包含在关系R中。个人同一性寓于R的独一无二的持有——处于一个现在的人和一个独一的未来人之间。既然对于个人同一性而言不止是关系R，那么我们能够合理地主张，就两者而言，是同一性重要。"

在回答这个诘难时，使用一些缩写将是大有帮助的。称个人同一性为PI，当某个关系独一无二地成立的时候，或者处于一种一对一的形式中的时候，称这个事实为U。我接受的观点能够以下面这个公式来陈述：

PI＝R＋U

我们绝大多数人深信PI重要，或曰具有价值。假定R也可能具有价值。那么有四种可能性：

（1）R没有U则没有价值。

（2）U强化R的价值，但是R即使没有U也具有价值。

（3）U对R的价值不造成任何差异。

（4）U减少R的价值（但是不足以消除这个价值，因为R＋U＝PI，后者具有价值）。

U的在场或缺场能够造成R的价值的巨大差异吗？正如我将论证的，这并不令人信服。如果我将与某个未来的人处于关系R中，U的在场或缺场对我与这个人的关系的固有性质没有造成任何变化。而且最为重要的必定是这个关系的固有性质。

既然如此，R没有U至少仍然会具有其绝大部分价值。加上U使R＝PI。如果加上U并不大幅地增加R的价值，R必定是从根本上重要的，而且PI最为重要就因为R的在场。如果U不给R的价值带来变化，PI重要只是因为R的在场。既然U造成小变化的主张能够令人信服，与R相比PI可能具有某种额外价值。但是这个价值要远远逊于R的固有价值。当U未能有的时候，PI的价值要远远少于R在PI缺场时会有的价值。

如果被独自提出的话，接受个人同一性并不重要这一观点会是困难的。但是我认为，当我们考虑那个分裂事例的时候，这个困难就消失了。当我们考虑为什么作为结果的人中没有哪一个是我的时候，我认为，经过再三考虑我们也能够看清这并不重要，或者只是有一点重要而已。

那个分裂事例支持还原论者的观点的一部分：即我们的同一性并不重要这一主张。但是这个事例并不支持还原论者的另一个主张：即我们的同一性能够是不确定的这一主张。如果我们舍弃同一性是重要的东西这一观点，我们就能够主张此处有一个对我的问题的回答。作

为结果的两个人中没有一个将是我。我将要死去。当我们认为同一性是重要的东西的时候，这个主张不可取地蕴含着，我应该把我的分裂当作几乎像通常的死亡一样糟。但是如果我们转而主张这种死亡方式大约像通常的幸存一样好，这种不可取性就消失了。

仍然存在着微小分歧的空间。尽管R是从根本上重要的，但是U能够造成一个小差别。我可能把我的分裂当作比通常的幸存更好些，或许当作比通常的幸存更糟些。

为什么我可能认为我的分裂比通常的幸存更糟些？我可能主张，我和每一个作为结果的人之间的关系不完全是在通常的幸存中重要的那个关系。这不是因为缺少某种东西，而是因为分裂引发出的东西太多。我可能认为每一个作为结果的人在某一方面具有比我的生活要糟的一种生活。每一个都不得不生活在这样一个世界中，其中至少一开始就有一个确切地像他自己的某个别人。这可能产生令人不快的离奇感。而且这也引发了实际难题。假设我最想做的是写书。这会是每一个作为结果的人都最想做的。但是他们两者都写这本书会是无意义的。两者都做他们最想做的事会是无意义的。

下面考虑一下作为结果的人和我所爱的那个女人之间的关系。我可以假定，既然她爱我，她将爱他们两者。但是她不能给他们两者以我们现在所给予他们每一个的那份专一的关爱。

在这些和其他方面，作为结果的人的生活可能大不如我的生活那样好。这可能证明我把分裂看作不像通常的幸存那么好是正当的，但它不能证明把分裂看作不那么好或像死亡那样糟是正当的。而且我们应该注意到，这个推理忽视了把这两个人的生活放在一起会是我的余生的两倍那么长这一事实。

代之以把分裂当作比通常的幸存要糟些，我或许视之为更好些。最简单的理由会是刚才给出的那个理由：我的余生时间翻倍。我或许有更多的特定理由。于是乎或许存在两个我非常强烈地想追求的终身生涯。我或许非常强烈地想要他们两者去做一个小说家和一个哲学

家。如果我分裂了，作为结果的人中的每一个就能从事其中的一个职业。而且彼此都乐见其成。就像我们能够对子女的成就感到自豪和快乐一样，作为结果，这两人中的每一个都会对另一个的成就感到自豪和快乐。

如果我有两个强烈但不相容的愿望，分裂就提供了一个实现两个愿望的方式，这个方式使每一个作为结果的人都高兴。以这种方式，分裂能比通常的存在还好，但是还有其他一些难题是分裂无法一揽子解决的。假设我在一个令人不快的职责和诱人的愿望之间遭受折磨。我不能通过虚拟打算让作为结果的人中的一个尽我的职责而另一个去满足我的愿望来彻底解决这个难题。我虚拟打算让他去尽我的职责的那个作为结果的人本身会处在职责和愿望之间的折磨之中。为什么那个去尽我那令人不快的职责的人会是他？我们能够预见到这里的麻烦。如果那个诱人的愿望不是由一个以上的人来实现的话，我的职责或许能够尽到。那个愿望或许是与某个只要一个伴侣的人私奔。那么这两个作为结果的人必定争相成为伴侣。那个在竞争中失败的人然后或许就不情愿地尽我的职责。尽管是以一种较不吸引人的方式，但是我的难题得以解决。

对于那些尚不相信还原论者的观点为真或同一性并不重要这一点的人们而言，这些评述将显得荒谬。这样的人可能会说："如果我不是作为结果的人中的任何一个，分裂就不会实现我的两个抱负。即便作为结果的人中有一个是一位成功的小说家，而另一个是位成功的哲学家，这也并未实现我的抱负中的任何一个。如果我的抱负之一是当一个成功的小说家，那么我的愿望是，我是一个成功的小说家。如果我停止存在而某个别人是一个成功的小说家，我的这个愿望将不会被实现。如果我将不是作为结果的人中的任何一个，那么所会发生的就是这样。"

这个诘难假定有这样一个真实的问题，即，我将是作为结果的人中的一个，还是另一个，抑或两个都不是？假定这些是三个不同的可

能性，其中的任何一个都可能会发生，这是自然而然的。但是就像我已经论证的，除非我是诸如笛卡儿的自我那样的一个独立存在的实体，这些不可能是三个不同的可能性。没有什么会使三者中任何一个可能会真实发生的这一点为真。（这与我有关存在一个对这个事例的最佳描述——我将不是作为结果的人中的任何一个——的主张相容。这一主张并不意味着我承诺了存在着不同的可能性这种观点。只有在一种情况下才会是这样，那就是在仅当其他描述中的一个可能一直为真的情况下，但是我否定了这种情况。）

我们能够给出一个不同的描述。我们可能会说我将是成为成功小说家的那个结果的人。但是，认为当且仅当我们称之为结果的人是我的时候我的愿望才会得以实现这点是错误的。我们如何选择表述这一事例没有什么理性的或者道德的重要性。

现在我将对我的主张加以回顾。当我讨论心理的、物理的和联合的谱系的时候，我论证说，我们的同一性可以是不确定的。这不是那个自然而然的观点。我们倾向于相信，对于"我将会死去吗？"这个问题必定总是有一个要么是要么否的答案。我们倾向于相信，我们的同一性必定是确定的。我论证过，除非我们是些像笛卡儿的自我那样的独立存在的实体，这不可能为真。我们不能既否定个人是这样的一个实体，又认定个人的连续存在具有与笛卡儿主义者们所归属于自我的存在的那些特别特征相同的特征。我退一步承认了我们可能一直是这样的一些实体。但是有相当的证据与这个观点相冲突。

如果我们否定我们是独立存在的实体，那么，就像我已经主张的，我们必须成为还原论者。还原论者的一个主张是，我们能够想象出这样的一些事例，其中"我将要死去吗？"这个问题没有什么答案，而且是空洞的。这看起来是有关那个联合谱系的主要事例的不可取性最小的一个观点。还原论者的另一个主张是，个人同一性并不是重要的东西。这看起来是有关那个我的分裂的事例的不可取性最小的观

点。就还原论者的这两个主张而言，第二个更加重要，因为它适用于我们自己的生活。

如果我们接受还原论者的观点，关于什么重要，我们还有进一步的问题要回答。这是些有关某些事实之理性的和道德的重要性的问题。但是，照还原论者的观点，所谓的"问题事例"就不再引发关于所发生的是什么的难题。即使当我们对于有关个人同一性的问题没有什么答案的时候，我们仍然能够知道所发生的一切。

我已经忽略了某个观点吗？我已经主张，如果我们拒斥有关我们是独立存在的实体的观点，我们应该接受某个版本的还原论者的观点。但是某些作者却辩护并不明显地属于这两者中的任何一者的一些观点。因而我将讨论这些作者们主张的是什么。

91.为何没有能满足两个令人信服的要求的同一性准则

除了在第83节中所讨论的论证之外，威廉斯针对心理准则提出了另一个论证。如果我事先陈述这个论证要表明什么可能再次会有所帮助。威廉斯主张，个人同一性准则必须满足两个要求。我将主张没有什么令人信服的个人同一性准则能够满足这两个要求。相反，依照还原论者的观点，那些类比要求却得到了满足。该论证从而赋予我们接受还原论者的观点的进一步的根据。但是威廉斯的论证并不假定还原论者的观点。因此，在讨论这个论证的时候，我将短暂地搁置这个观点。它在舞台边等候，待情节需要的时候重新登场。

威廉斯的论证发展了雷德的一个评述，这个评述针对的是洛克有关谁"具有对现在的和过去的活动的意识，它们就属于谁"的主张。正如雷德所写到的，这蕴含着，"如果同一个意识能够被从一个智慧生物转移给另一个……那么两个或者二十个智慧生物就可能是同一个人了"。[12]

威廉斯的论证如下。同一性在逻辑上是一种一对一的关系。一个人与一个以上的人同一在逻辑上是不可能的。我与两个不同的人不可

能是同一个人。正如我们已经见到的，心理的连续性在逻辑上不是一对一的关系。未来的两个不同的人两者都能在心理上与我相连续。既然这两个不同的人不能两者都是我，心理的连续性不能是同一性的准则。威廉斯然后主张，一个同一性的准则要成为可接受的准则，本身在逻辑上必须是一对一的关系。它一定不可能是一个人和两个未来的人之间的关系。他因而主张同一性的准则不能是心理连续性。[13]

一些人回应说，这个准则或许诉诸非分支心理连续性。我已经讨论过的正是这个准则的这个版本。照我所称的心理准则，如果一个未来的人将与我处于关系R中，而且没有他人将与我处于关系R中，这个未来的人将是我。鉴于这个准则的这个版本在逻辑上是一对一的关系，对于它回答了威廉斯的诘难这一点已有所主张。[14]

威廉斯拒斥这个回答。他主张

> 要求（1）：一个未来的人是否是我，必定只依凭我们之间关系的固有特征。不能依凭发生在其他人身上的是什么。
>
> 要求（2）：既然个人同一性具有巨大的意义，同一性是否成立不能依凭一个微不足道的事实。[15]

这些要求两者都是令人信服的。而且其中没有哪一个要求是由非分支的心理连续性所满足的。因此，威廉斯拒斥心理准则的这个版本。

这个诘难看起来过于抽象，难以令人信服。如果我变更一下我以之开始的那个想象出的故事，其力量就能够得到显示。考虑一下简单电子传输，其中扫描仪毁坏了我的大脑和躯体。在我的图像被用电子束传到火星之后，复制器创造出一个完美的有机复本。火星上的我的复制品将认为他是我，而且他将在心理上的每一个方面都与我是相连续的。

假设我接受那个诉诸一对一形式的关系R的心理准则。而且假设我们接受广义的版本，这个版本允许R具有任何可靠的原因。这个准

则蕴含着火星上的我的复制品将是我。但是我们或许得悉我的图像还被用电子束传到了木星的一个卫星木卫一上。然后我们必定主张在火星上苏醒过来的将是我，而且如果我的图像被木卫一上的科学家们所忽视的话，我将继续存在。但是如果在木卫一上科学家们后来造出了我的另一个复制品，当这个复制品苏醒过来时我将不再存在。尽管火星上我周围的人将不会注意到任何变化，但是在那个时刻一个新人将在我的大脑和躯体中出现。威廉斯会诘难说，如果我的确在火星上苏醒过来，我是否在那里继续存在，正如我们所主张的，不能依凭在数百万英里之遥的木星附近的某个别人身上发生什么。我们的主张违背了要求（1）。

正如我已经论证的，根本上重要的是我是否将至少与一个未来人处于关系R中。我是否还将与某个他人处于关系R中相对来说是微不足道的。照心理准则的这个版本，我是否将与某个未来人同一有赖于这个相对微不足道的事实。这违背要求（2）。

威廉斯会添加上这些评述。一旦我们看到电子传输会产生我的许多复制品，他们是彼此不同的人，我们就应该否定即使他们只造了一个单一的复制品我也会事实上在火星上苏醒过来这一点。如果他们制造了两个复制品，它们不会两个都是我。如果它们不会两个都是我，但是它们却是以同一方式制造的，我们应该断定两个都不是我。但是我与诸复制品中的一个的关系是内在地相同的，无论他们是否还造了另一个。既然同一性必定依赖于一个关系的内在特征，我就会不是两者中任何一个复制品，即便他们并未制造另一个。[16]

威廉斯用这个论证来支持物理准则的一个非还原论者的版本。（这个版本之所以是非还原论者的，是因为它假定个人同一性是一个进一步的事实，这个事实要求物理的连续性而不是寓于物理的连续性。）正如威廉斯所承认的，一个类似的论证能够挑战这个观点。他拒斥心理准则，因为它诉诸这样的一个关系，这个关系可能以分支形式出现在一个人与两个、甚至更多的未来人之间，心理准则因而不能

满足他的两个要求。然后他考虑了有关他的物理准则版本也不能满足他的两个要求的诘难。物理连续性能以一种分支形式呈现出来。正如他所写到的，"想象一个人像变形虫一样分裂成他自己的两个物像是可能的。"[17]

威廉斯对这个诘难给出了两个回答。假设我们认为我的大脑和躯体在物理上与一个被我的双亲视为己出的人的大脑和躯体相连续。我们很想知道这个物理连续性是否呈现为一种反常的、分支形式。如果我们知晓这个物理上连续的大脑和躯体的整个历史，这会"不可避免地揭示"是否有过一个像阿米巴（变形虫）一样分裂的事例。在心理连续性事例中这个可比的主张不为真。我们或许知道在地球上的我与在火星上的我的复制品之间的心理连续的整个历史，但仍不能知道我在木卫一上有另一个复制品。分叉对物理准则和心理准则而言都是一个难题。但是对物理准则而言问题的严重性要小一些，因为获悉难题什么时候出现在原则上会更加容易一些。

威廉斯还主张，当一个物理对象分裂的时候，这是其时空连续性的一个内在特征。相反，当两个人在心理上与一个早先的人相连续的时候，这个事实不是这些关系中任何一个的内在特征。与心理准则不同，物理准则满足要求（1）。

我的想象出的分裂事例对物理准则提出了几个诘难。我曾从两个方面修正了这个准则。我首先考虑了我的大脑被移植到我的那个同卵双胞胎兄弟的躯体的事例。经再三思考，显然在这个事例中我是那个幸存的接受者，而不是那个死亡的捐赠者。如果我的大脑被赋予一个新的躯体，这正是获得一个新心脏、一个新肺脏等等情况的极限事例。物理准则只应该诉诸我的大脑的连续性。然后我诉诸很多两个半脑中的一个半脑被毁的人幸存下来的事实。既然这些人幸存了下来这件事是清楚明白的，物理准则就应该诉诸这种连续性，但不是整个大脑的连续性，而是足够支持有意识生活的那部分大脑的连续性。

这样的连续性不是逻辑上的一对一关系。在那个想象出的我分裂的事例中，两个作为结果的人中的每一个都具有足以支持有意识生活的我的部分大脑。我们不能以它永远不会发生这一主张来摒弃这个事例。其最麻烦的特征——意识分裂，已经发生了。分裂下脑可能仍旧是不可能的。但这是一个纯技术的不可能性。同样，电子传输可能永远不可能。但是这样的不可能性并未削弱威廉斯针对广义心理准则所作的论证。而且如果他在那个论证中诉诸这样的一些事例，他也不能摒弃那个想象出的我完全分裂的事例。

威廉斯的论证所蕴含的是，在这个事例中我将不再存在，而且作为结果的人两者都会是新的人。从而他必须修正物理准则，以便它呈现为一个非分支的形式。有人或许诉诸我描述过的那个版本。此乃

物理准则：如果有一个未来的人具有我的大脑中足以成为一个活生生的人的大脑的足量部分，那么这个人将是我，除非还将有某个别人具有我的大脑的足够部分。

威廉斯会拒斥这个准则，因为他的要求中的两条它全违背了。

再次值得举例说明。假设我的分裂进程如下。我有两个大脑受了致命伤的兄弟，杰克和比尔。外科医生首先摘取并分割了我的大脑。然后这两个半脑被分别送到医院两厢的手术室内，它们在那里将被移植到我的两个兄弟的体内。如果我们诉诸物理准则，我们必须主张如下。假设我的大脑的一半被成功地移植到杰克的体内。在被移植之前，另一半却无意间掉落到水泥地面上。如果这是所发生的，我将在杰克的躯体中苏醒过来。但是如果另一半被成功地移植的话，我将不会在两者中的任何一个躯体中苏醒过来。

这些主张违背了要求（1）。我是否是杰克躯体内的那个人应该只依赖我与这个人之间的关系的内在特征。认为依赖于医院的另一侧发生什么不会是令人信服的。其他地方无论发生什么看来就像木卫一上的科学家们无论是否制造一个我的复制品一样无关。无论比尔身上发生了什么，也无论我的大脑的另一半发生了什么，我与那个在杰克

躯体内的人的关系必定是同一不二的关系。这个主张受到物理准则的否定。而且，与我的半个大脑将在杰克的躯体中幸存下来这个事实的重要性相比，对我来说另一半个大脑发生什么就是相对微不足道的。因此这个准则也违背要求（2）。

威廉斯可能提出

> 新的物理准则：一个未来的人将是我，当且仅当这个人既是活生生的人又具有我的大脑的一半以上。[18]

这个关系的一个内在特征是，它只呈现为一对一的形式。对两个未来的人而言，两者都具有我的一半以上的大脑在逻辑上是不可能的。这个准则从而满足要求（1）。

不过，它未能满足另一个要求。我既会在某个未来的人具有我的一半大脑的时候与之在心理上完全相连续，又会在某个未来的人具有我一半大脑多一点的时候与之在心理上完全相连续。而且对那些相信重要的是物理连续性的人们而言，这两个情况之间的差异一定是微不足道的。第二种情况只牵涉到稍微多一点的细胞的连续性。某个未来的人是具有我的一半大脑还是具有一半多一点的大脑是个微不足道的事实。新的物理准则从而违背要求（2）。

对这个准则还有另一个诘难。有人或许受了这样的伤，以至于导致他一半以上的大脑失去功能。尽管这样的人的大半个身体会瘫痪，而且或许需要置于一个心肺机器中，他的精神生活会是不受影响的。一个大脑的一小半就会足以提供完全的心理连续性。我们会自然地认为这样的一个人在受伤之后幸存下来。但是，照新的物理准则，我们必须主张这样的一个人不再存在。在他的躯体中的那个人是某个别人，一个仅确切地像他的一个新人。这是难以置信的。这是对这个准则的第二强烈诘难。

在其所有的版本中，物理准则都面对一些强烈的诘难。而且对心理准则有相似的诘难。威廉斯的两个要求都是令人信服的。我们已经

发现，没有什么可信之处的同一性准则能够两个要求都满足。（如果我们是像笛卡儿的自我那样独立存在的实体的话，我们的准则或许满足这些要求；但是我们具有充分的理由拒斥这个观点。）

现在回到还原论者的观点。重新考虑一下那个有关我的半个大脑被成功地移植到杰克体内的事例。我与那个在杰克体内苏醒过来的人的关系是什么？这个关系是心理连续性，带有其正常原因，即我的大脑的足够部分的连续存在。此外还有非常接近的物理相似性。作为一个还原论者，我主张我与杰克体内的那个人的关系包含着从根本上说重要的东西。无论在他处的他人身上发生什么，这个主张都有效。凭借一处修正，我的观点满足威廉斯的第一个要求。他主张我是否将是某个未来的人应该只依赖于我与这个未来的人的关系。我持一种相似的主张。我不是问我是否将是某个未来的人，而是问我与这个人的关系是否包含重要的东西。像威廉斯一样，我能够主张，该问题的答案必须只依赖于我与这个未来人之间关系的内在特征。

还原论者的观点能够满足要求（1）的这个修正版本。假设另一个手术取得成功。有人在比尔体内苏醒过来。照我的观点，这并不改变我与杰克体内的那个人之间的关系。它充其量对这个关系的重要性造成一点影响。这个关系仍然包含从根本上说重要的东西。既然这个关系现在处于一种分支的形式，我们被迫更改其名称。我们不能够把这个关系的每个分支称为个人同一性。但是这个关系名称的更改没有什么重要意义。

这个还原论者的观点也满足要求（2）的比拟。个人同一性的判断具有巨大的重要性。因此，威廉斯主张，如果我们的根据中没有什么重要的不同之处，我们就不应该作出一个这样的判断而否定另一个判断。照此还原论者的观点，我们应该接受我们所给予同一性判断的重要性，而且应该把这个重要性给予一个不同的关系。照此观点，重要的是关系R：带有那类确当原因的心理的联系性和／或心理的连续性。与同一性不同，这个关系不因事实中的一个微不足道的差异而不

能成立。如果这个关系不能成立，那么在事实中就有一个深刻的差异。这满足要求（2）。

在我分裂的那个事例中，尽管我与作为分裂的人中的每一个人的关系不能称之为同一性，但是它包含从根本上说重要的东西。当我们否定此处的同一性的时候，我们用不着否定一个重要的判断。既然我与作为结果的人中的每一个人的关系宛如同一性，它可能承载同一性的绝大多数通常的蕴含。于是就有可能主张说，即使因为另一个移植取得成功，杰克躯体中的那个人不能被称为我，但是他同样能因为我所做的而应受赏罚。比尔躯体中的那个人也是如此。正如威金斯所写的："一个不法之徒通过谋划他自己的分裂几乎不会逃脱惩罚。"[19]

此处有些问题有待回答。如果不法之徒被判20年监禁，作为分裂结果的每一个人都服刑20年呢，还是只服刑10年？我在第十五章中对这些问题中的一些问题进行了讨论。这些问题并不会使我已经做出的一般性主张投下一层怀疑的阴影。如果我们接受还原论者的观点，所重要的正是R而不是同一性。

人们可能想，如果是这样，我们应该把我们现在给予个人同一性的重要性给予R。结论并非如此。如果我们认为个人同一性具有巨大的重要性，这可能是因为我们相信非还原论者的观点。如果我们改变我们的观点，而且成为还原论者，我们也可能改变我们有关个人同一性的重要性的观点。我们可能接受这一点：即关系R几乎具有，按照还原论者的观点，个人同一性所具有的所有重要性。而且我们可能接受，照此观点，从根本上说重要的东西不是个人同一性而是关系R。但是我们会认为，如果非还原论者的观点为真的话，这些关系在其重要性上两者都比个人同一性会有的重要性要小得多。在第十四章和第十五章我对这个信念进行讨论。

这个信念并不影响我有关威廉斯的要求所做的主张。如果我们假定同一性是重要的，我们不能满足这些要求。既然我们应该拒斥非还原论者的观点，那么我们的同一性准则就既不是物理准则又不是心理

准则。而且，正如我已经论证的，两个准则的任何一个都没有满足威廉斯的两个要求的令人信服的版本。

我添加这些评述。既然我们已经看到同一性并不是很重要，我们就不应该尝试修正或者扩充我们的同一性准则，以便它更加经常地与重要的相重合。按照对于个人同一性的任何一种自然的理解，这样的一个重合只会是部分的，就像那个分裂事例所表明的那样。而且修正我们的准则本身可能令人误解地提示着同一性是重要的。

对于要求一个同一性的准则在逻辑上是一对一的，而且方式并非是随意的，威廉斯还给出了一个别的根据。"除非有某个这样的要求，我不清楚一个人如何维护和解释这样一个明显的真理，即同一性概念和确切相似概念是两个不同的概念。"[20]

还原论者的观点保留和解释了这个真理。我已经描述了一些事例，在这些事例中有两个人确切相似但是并非在量上同一。在那个我与我的复制品交谈的支线事例中，就可能如此。我们因而理解"他与我是同一个人还是只与我确切相似的另一个人？"这一问题。我已经主张过，在某些事例下，诸如在物理谱系中央的那些事例中，在作为结果的人是我和他是某个只与我确切相似的别人之间没有真正的差异。在某些事例中，还原论者的观点的确蕴含着，数的同一性和确切相似性之间不存在真正的区别。但是鉴于它承认在别的一些事例中存在这样的区别，它就保留和解释了这些是不同的概念这一真理。

我之所以对威廉斯的诸般主张进行讨论，想弄清它们是否提供某个有别于还原论者的观点和那个认为我们是独立存在的实体的观点这两者的第三种观点。我断定这些主张并不提供这样一个第三种观点。它们提供了接受还原论者观点的一些进一步的根据。威廉斯的两个要求都是令人信服的。如果我们认为同一性是重要的，我们就不能满足这些要求。但如果我们接受还原论者的观点，而且诉诸关系R，则我们就能够满足这些类似的要求。

92.维特根斯坦和佛陀

维特根斯坦会拒斥还原论者的观点。他认为我们的概念依赖于某些事实的成立，而且我们不应该考虑其中这些事实不再成立的那些想象的事例。为还原论者的观点进行辩护的那些论证所诉诸的就是这样的一些事例。

这个分歧只是部分意义上的。绝大多数人具有有关这些想象性的事例的信念。正如我已经论证的，这些信念蕴含着我们是些与我们的大脑和躯体相分离的独立存在的实体，是些其存在要么全有要么全无的实体。还原论者的主要主张是我们应该拒斥这些信念。维特根斯坦本来也会同意的。鉴于在这个主张方面意见一致，我无需讨论维特根斯坦的观点，或者某个其他的相似观点，诸如威金斯所提出的观点等等。[21]

除了我将很快提到的两个例外之外，我认为我现在已经考虑了那些在这个辩论中需要加以考虑的观点。我可能不知道其他某个发表的观点。而且我尚未考虑在不同的时代或者文明中所持的观点。这个事实提示着一个令人担忧的可能性。我认为我的诸主张适用于所有时代的所有的人。发现它们只是现代欧美文化中的一个思想路线的组成部分会是令人担忧的。

所幸的是，情况不是如此。我主张，当我们问个人是什么以及他们如何连续存在的时候，根本的问题在于从两个观点间作出选择。照其中的一个观点，我们是些与我们的大脑和躯体相分离的独立存在的实体，是些其存在要么全有要么全无的实体。其中的另一个观点是还原论者的观点。而且我主张，就这些观点而言，第二个观点为真。正如本书附录部分所表明的，佛陀本来也会同意。还原论者的观点不仅仅是一个文化传统中的一部分。正如我所主张的，它可能是关于所有时代的所有的人的真实观点。

93.我实质上是我的大脑吗?

内格尔提出的一类观点是我还没有讨论过的。他提出，我的心理连续性的原因是什么，我实际上就是什么。倘若就我们现在所知道的而言，实际上我就是我的大脑。此外，照此观点，我实质上是我的大脑。我不会决定接受一个与之不同的关于我自己的观点。

内格尔以三种方式支持这个观点。他给出了两个论证，对此我将在附录D中尝试给予回答。他还主张说，他的观点是直觉地令人信服的。就像他所写的，他的大脑

> 在我看来是某种离开它我就不会幸存的东西——所以，如果一个物理上明显有别的我的复制品被生产出来，这个复制品在心理上与我相连续，虽然我的大脑已经被毁，它就不会是我，而且其幸存就不会好得像我的幸存一样。

内格尔在此处考虑的是一个类似电子传输的事例。扫描复制器会在这里的地球上使用。我会把它当作一个穿越纽约曼哈顿的更为安全的方式来使用，或者把它当作在走过曼哈顿的过程中遭受致命刀伤时的一个治疗方案。作为一种预防措施，在每次走路之前，我都会让人储存起我的一个新图像。

内格尔提出，电子传输不仅不像通常的幸存一样好，而且几乎像死亡一样糟。正如在描述了他所想象出的事例之后他所写到的：

> 我将不会在那个夜晚幸存下来……那个复制品不是我。努力鼓起勇气，我为末路做好准备。

照内格尔的观点，这表明，个人同一性是重要的。我承认，在一个像电子传输这样的事例中，很多人会接受内格尔的观点。他们会认为重要的是他们的大脑的幸存。在附录D中我描述了这点在其中更加难以置信的两个事例。

既然内格尔的观点在某些情况下是直觉地令人信服的，而且附录D可能不能回答内格尔的论证，我将在第98节中对这个观点和类似的一些观点给出一种不同的回答。

还有另一个观点我应该予以考虑。我在附录E中解释，何以诺齐克的观点是还原论的一个版本，尽管两者有明显的不一致。

94.真的观点可信吗?

内格尔一度主张，即使还原论者的观点为真，对我们来说在心理上相信这一点也是不可能的。因而我将简短地回顾我在上文所给出的论证。然后我将询问我是否能够诚实地主张相信我的诸结论。如果我能够，那么我将认为我不是独一无二的。至少会有某些别人能够相信真理。

我将首先厘定我已经表述了真的观点这一主张。要准确解释一个还原论者主张什么是困难的。要准确解释时间中的同一性涉及什么也是如此。我在表述还原论者关于同一性的观点的时候，错误很可能在所难免。

这样的错误不会整个削弱我的那些论证的根基。维特根斯坦曾经提出这样一个比拟。假设我正在重新排列图书馆中的藏书。在重新排列图书之初，我把两本书置于某个特定的书架上，因为这些书籍应该归放在一起。我可能知道，在这项重新排列工作大功告成时这两本书将出现在一个与之不同的书架上。但是现在仍然值得把它们归放在一起。如果它们应该在一起，在这次重新排列之后它们将仍然在一起。

我主张个人不像笛卡儿的自我——其存在要么全有要么全无的一个在者。个人就像一个国家。在那个对时间中的同一性的真实说明中，这两类实体如影随形。它们就像我起初放回到那个特定书架上的那两本书一样。在我有关还原论和同一性的诸主张中，差错或许难免。这也许像那两本图书应该归放到另一个不同书架上这一事实一样。但是我的主要主张是，个人像国家，而不是像笛卡儿的自我。如果这个主张为真，就不会被我的错误所削弱。在所谓毫无差错的说明中，个人和国家仍然会走到一起。在对它们在时间中的同一性的说明上，就其实质性特征而言会是相似的。

我区分出关于个人本质的两个不同观点。照非还原论者的观点，一个个人是一个与其大脑、躯体和经验相分离的独立存在的实体。照其人所共知的那个版本，一个人是一个笛卡儿的自我。照我所捍卫的还原论者的观点，个人存在着。而且一个个人与其大脑和躯体，以及其经验相分离。但是个人不是独立存在的实体。一个个人在任何时期的存在就寓于其大脑、躯体、思想的思维、行动的实施以及许多其他的物理和精神事件的发生之中。

鉴于这些观点在有关个人的本质方面意见相左，它们在关于时间中个人同一性的本质方面也是意见不合。照还原论者的观点，个人同一性只不过涉及物理的和心理的连续性。正如我所论证的，这两者都能够以一种非人称的方式加以描述。这两类连续性都能够不通过主张经验为某个个人所拥有而得到描述。一个还原论者还主张，个人同一性不是重要的。个人的同一性仅仅涉及到某些种类的联系性和连续性，要是这些保持一对一的形式的话。这些关系正是重要的。

照非还原论者的观点，个人同一性是重要的。而且它并不只是涉及物理的和心理的连续性。它是一个独立的进一步事实，在任何情况下它必定要么完全成立，要么完全不成立。心理的统一性是由所有权来解释的。任何时候的意识的统一性都是由数个经验正被某一个人所具有这点来解释的。一个个人的生活的统一性也是以相同的方式来解释的。这数个主张，正如我已经论证过的，必定要么一起成立要么一起不成立。

我退一步承认非还原论者的观点可能本来为真。例如，可能本来有证据支持肉体复活的信念。但是事实上没有充分的证据有利于这个观点，反而有相当的证据不利于它。

有些证据是由那些头脑分裂的人的实例提供的。由于他们的两个半脑被隔断了联系，几个人都具有两个意识流，在其中的每一个中他们都没有察觉到另一个。我们可能主张，在这样的情况下，同一个躯体中有两个不同的人。这就像把这样一些实例当作下文回顾的我的分

裂的那个想象事例一样对待。我们的备选方案是，就这些实例而言，主张存在着具有两个意识流的单独的一个人。

如果我们作出这样的主张，我们如何能够解释每一个意识流中的意识统一性？我们不能通过主张每一个意识流中的各色经验正在被同一个人或者经验主体所具有来解释这个统一性。这就把两个意识流宛如当作一个来描述。如果我们认为意识的统一性必须通过把不同的经验归属于一个特定的主体来解释的话，我们必须主张尽管在这些事例中只有一个单独的人，但是有两个经验主体。因此我们必须主张，在一个人的生命中能够有不是个人的一些经验主体。难以相信真的存在这样一些经验主体。这些事例通过还原论者的心理准则能够得到更好的解释。这个准则主张，任何时刻在一个意识流中都有一个对那些经验的感知状态，在另一个意识流中又有另一个对那些经验的感知状态。

尽管这些头脑分裂的事例为非还原论者的观点引出了这个难题，但是这些事例只不过是不利于这个观点的证据中的一小部分而已。没有什么证据表明心理连续性的承载者是某种像笛卡儿的自我那样的其存在要么全有要么全无的东西。而且有相当的证据表明这个连续性的承载者是大脑。有相当的证据表明我们的心理特征依赖于我们大脑中的状态和事件。一个大脑的连续存在不一定要么是全有要么是全无。物理的联系性可以是一个程度问题。而且在数不尽的实例中，心理的联系性只是以某些方式维持着，抑或在某种弱化的程度上维持着。

我们有充分的证据拒斥非还原论者的观点。我主张，还原论者的观点是惟一的备选。我考虑过第三种观点的可能性，但是没有发现什么既是非还原论者的又是有充分理由予以接受的观点。更确切些说，尽管这些其他观点在其他方面有别，但是这些令人信服的观点并不否定还原论者的主要主张。它们认同：我们不是一些独立存在的实体，即不是一些有别于我们的大脑、躯体和经验、其存在必定要么全有要么全无的实体。

除了作出这些有关事实的主张之外，我还作出了一些有关我们的信念的主张。我们得悉，当我们考虑某些想象事例的时候我们具有这些信念。

那个医生通过擅改我的大脑渐渐地清除我的所有心理连续性的事例就是这诸多事例中的一例。我还表述过其他三个事例域。在心理谱系和物理谱系中，在我和某个未来的人之间不是会有心理联系性的一切可能程度，就是会有物理联系性的一切可能程度。在联合谱系中，会有这两类联系性的一切可能程度。

像我们中的绝大多数人一样，我强烈地倾向于认为未来的人要么是我要么是某个别人。而且我倾向于认为在这两个结果之间总是有深刻的区别。既然我不是一个独立存在的实体，这些信念就不能为真。只要考虑一下联合谱系，这一点就会得到最好的说明。在那个近端事例中，在我身上不会发生任何事情，作为结果的人当然会是我。在那个远端事例中，其中在我和作为结果的人之间不会有什么联系，这个人当然是某个别人。如果任何一个未来的人必定要么是我要么是某个别人的话，那么在这个事例域中必定有一条分界线，在这个界线之前作为结果的那个人会是我，超过了这个界线他就会是某个别人。如果在某个人是我与他是某个别人之间总是存在深刻的差异的话，那么在这个事例域中的这两个事例之间必定有深刻的差异。尽管我们永远不会发现这个差异来自何处，但是必定会有这样的一个深刻差异。这些主张是虚妄的。在这个事例域中的任何相邻事例之间没有什么深刻差异。惟一的差异是，在这些事例域的一个事例中，外科医生将置换稍多一些的细胞，而且会造成一个更小的心理变化。

当我考虑这个事例域的时候，这迫使我舍弃了上述两个信念中的至少一个信念。我不能继续既认为任何未来的人要么是我要么是某个别人，又认为在这两个结果之间总是有深刻的差异。

这还迫使我接受还原论者的观点的另一部分。假设我在这个谱系的中间地带将要经受这些手术中的一个手术。我知道，在我和作为结

果的人之间将有某些类型的和程度的物理的和心理的联系性。照还原论者的观点，在认知这些事实时我获悉有关什么将发生的整个真相。当我将要失去意识时，我可能问，“我将要死去呢，还是我将是那个作为结果的人呢？”而且我倾向于相信这些总是两个不同的可能性，其中的一个必定是真相。照还原论者的观点，此处这是一个空洞的问题。有时在某个未来的人是我与他是某个别人之间存在一个真正的差异。但是在联合谱系的中间地带的那些事例中没有这样的真正差异。差异会是什么？是什么能使作为结果的人要么会是我要么会是某个别人为真？既然我不是一个独立存在的实体，就没有什么会使这些成为不同的可能性，其中的任何一个都或许为真。在这些事例中，我们会说作为结果的人将是我，抑或会说我将死去而他将是某个别人。但是这里这些并不是不同的结果。他们只是对同一结果的不同描述。

为了例证这些主张，我转述了休谟所作的比较。个人就像国家、俱乐部或者政党一样。如果我们正在考虑个人之外的这些其他实体的话，我们绝大多数人接受一种还原论者的观点。谨记那个政党分裂成两个竞争对手的事例。我们能够问，“原来的那个政党不再存在了，还是它作为分裂后的两个政党中的这一个或者那一个而继续存在？”但是我们并不认为这是关于不同可能性的——其中的一个必定是所发生的——一个真实的问题。这个问题是空洞的。纵然我们没有对这个问题的答案，但是我们会知道关于所发生的是什么的整个真相。

既然我们采信还原论者关于政党、俱乐部或者国家的观点，我们就大致理解关于个人还原论者的观点在主张些什么。但是我们绝大多数人强烈倾向于拒斥这个观点。我们强烈地倾向于相信在某个未来的人是我与他是某个别人之间总是有差异。只考虑联合谱系可能不足以说服大家成为还原论者。因此我给出了一些进一步的论证。

其中的一个论证诉诸于想象的我的分裂事例。一个意识流的分裂或许被主张为深刻地不可能。但是所发生的必定是可能的：在数个人

的生命中这确实已经发生了。我的想象的分裂事例是这些实例的自然延伸。

在这个想象的事例中，我的大脑的每一半大脑都被成功地移植到另外的躯体内。在我身上发生了什么？除非我们荒诞地曲解个人概念，否则可能的答案唯有：我将是作为结果的人中的这一个，或者是那一个，抑或不是其中的任何一个。如果我们认为同一性是重要的，这些答案中的每一个都是难以接受的。有鉴于两个作为结果的人的确切相似性，难以相信我将是这两个人中的一个。如果我两者都不是，而且同一性是重要的，我应该把分裂视作等同于死亡。但这也是难以置信的。我与每一个作为结果的人之间的关系包含着幸存所需的一切要素。这个关系不能称为同一性，因为而且仅仅因为它处于我与两个未来的人之间。在通常的死亡中，这个关系处于我与根本没有的未来人之间。尽管双重幸存不能以同一性语言加以表述，但是它并不等同于死亡。二并不等于零。

这个想象的事例支持还原论者的观点的另外部分。不仅难以相信每一个可能的答案，而且难以看清该事例如何涉及到其中的任何一个都可能为真的不同的可能性。如果我不是一个笛卡儿的自我，使我会是两个作为结果的人中的这一个或者那一个这点为真的是什么？如果这些是不同的可能性，那么差异何在？对这个问题好像没有答案。每一个作为结果的人都将有我的半个大脑，而且都将与我在心理上相连续。我们看来被迫断定这是对该事例的一个完全描述。我们理解"我将是这两个人中的这一个呢，还是那一个，抑或哪一个都不是"这个问题，但这又是一个空洞的问题。在此这些不是其中的一个必定为真的一些不同的可能性。这些只是对同一结果的不同描述。

最佳的描述是，我不是作为结果的两个人中的任何一个。但是这并不蕴含我应该把分裂当作像死亡一样糟。正如我所论证的，我应该把这看作大约像通常的幸存一样好。对某些人而言，它会比幸存略好；对另一些人而言，它会比幸存略糟。既然我与两个作为结果的人

不能是同一个人，但是我与他们每一个人之间的关系包含着在通常的幸存中从根本上说重要的东西，那么该事例表明同一性无关紧要。重要的是关系R：带有确当种类原因的心理联系性和／或者心理连续性。

现在我已经回顾完了支持还原论者的观点的主要论证。我发现相信这个观点是不可能的吗？

我的发现是这样的。我能够在理智的层面上或者反思的层面上相信这个观点。我折服于赞同这个观点的那些论证。但是我认为情况很可能是，在某个别的层面上，我将总是有疑问。

当我考虑这些想象出来的某些事例的时候，我的信念最为坚定。令我信服的是，如果我分裂，有关我是作为结果中的这一个人、还是另一个人抑或两个人都不是的问题是一个空洞的问题。我相信没有什么会使这些成为不同的可能性——其中的任何一个都可能是真实地发生的。而且令人信服的是，在第三谱系的中间地带的那些事例中，有关作为结果的那个人是否是我的问题是一个空洞的问题。

当我考虑某些其他事例的时候，我的信念就较不坚定了。电子传输就是这样一个事例。我想象我在舱室中，将要按下那个绿色的按钮。我可能突然间产生怀疑。我可能试探改变主意，宁愿付出更大一笔费用乘坐太空船旅行。

我猜想，回顾我的论证决不会整个地消除我的怀疑。在反思的或者理智的层面上，我会仍然坚信还原论者的观点为真。但是在某个较低的层面上，我会仍然倾向于认为在某个未来的人是我与他是某个别人之间必定总有一个真实的差异。当我在摩天大楼顶层临窗观景的时候，类似的情况亦然。我知道我并不处在危险之中。但是从这个令人头晕目眩的高度俯视，我感到害怕。如果我将要按下那个绿色按钮，我会有类似的非理性的恐惧。

添加这些评述可能有所帮助。照还原论者的观点，我的继续存在只涉及物理的和心理的连续性。照非还原论者的观点，它涉及一个进

一步的事实。相信这个进一步的事实，相信与那些连续性相比这是一个深刻的事实和真正重要的事实，这些是自然的。当我担心在电子传输中我将不会抵达火星的时候，我的担心是那个非常的原因不能引起这个进一步的事实。正如我已经论证的，没有这样的事实。我所担心的将不会发生，永不发生。我想要这个在火星上的人以这样一种特别密切的方式像我，其中将永远没有什么未来的人是我。我的连续存在决不牵扯这个深刻的进一步事实。我担心将缺少的东西总是缺少的。即使一趟太空飞船旅行也不会产生那个我倾向于相信的进一步事实。

当我最终看清我的连续存在并不涉及这个进一步事实的时候，我就丧失了选择太空飞船旅行的理由。但是从我早先的信念立场来判断，这不是因为电子传输大约像通常的幸存一样好，而是因为通常的幸存大约像电子传输一样糟，或者比电子传输略好一点。通常的幸存大约像被摧毁和被复制一样糟。

通过对像这样的一些论证的预演，我或许足以缓解我的担心。我或许能够使自己按下那个绿色的按钮。但是我预期，我永远不会完全失去我对非还原论者的观点的直觉信念。要泰然地相信我的还原论者的结论有一定的难度。相信个人的同一性并不重要这一点是困难的。如果明天某个人将处于剧痛中，难以相信这个剧痛是否被我感觉到的问题是一个空洞的问题。而且难以令人相信，如果我将要失去意识，“我将要死去吗？”这个问题会没有什么答案。

内格尔曾经主张，认为还原论者的观点从心理上说是不可能的。佛陀主张，尽管这很难，却是可能的。我发现佛陀的主张为真。经过回顾我的论证，我发现，尽管在反思的或者智力的层面上认为还原论者的观点是困难的，却是可能的。我那些剩余的怀疑或者说担心在我看来是非理性的。既然我能够相信这个观点，我假定其他人也能够这么做。我们能够相信有关我们自己的真相。

---

[1]Sperry，第299页。

[2]Nagel（5），重印于Nagel（4），第152页。

[3]Nagel（5），重印于Nagel（4），第153页。

[4]Madell，第137页提出，使我的经验之为我的经验的东西不是它们被一个特定的经验主体——我——所拥有，而是它们具有作为我的经验的那个属性。按照这个观点，精神空间的位置是由大量的不同属性的存在所给定的，每一存活的人一个。我同意马代尔的观点，就是我和他会有两个质上相同一的同时发生的经验，但又明显地不同。但这无需是因为那些经验之一具有是我的经验的那个独一无二的属性，而另一个则具有是马代尔的经验的那个独一无二的属性。它可能只是因为这些经验之一是这个经验，发生在这个特定的精神生活之中，而另一个是那个经验，发生在那个别的特定的精神生活之中。这两个精神生活或许得通过它们与一对不同的人的躯体的联系而加以公共指称。（那些头脑分裂实例提供了对马代尔观点的诘难。把是我的（being mine）作为一个属性（property）来对待看来并不令人信服。把不同的特定事物或者事件区分开来的东西不是各自具有一个独一无二的属性。物理事物或者事件能够因处于不同的地点而彼此不同。当我想到我当前的思想是我的思想的时候，我并非通过指称一个空间位置而验明这个思想。我的验明性的指称实质上包含一个指标性的（indexical）词，或者一个演示性的（demonstrative）词，而不是对一个独一无二的属性的归属。我能够使用指标性的"这个"或者指标性的"我的"。我的主张是，既然我能够使用"这个"的自指用法，我就不需要使用"我的"。）

[5]一些人否认"我"能以使用"这个"的自指用法的方式得到解释。至于一个支持我所采纳的观点的论证，请参见Russell。

[6]我沿袭Shoemaker（1），第22页。

[7]Quinton，见Perry（1），第60页。

[8]Wiggins（1），第50页。我之所以决定学哲学几乎完全是因为我受到威金斯所想象的事例的吸引。

[9]参见Wiggins（1），第40页。我把这个所提及的说话方式及对它的诘难归功于迈克尔·伍兹。

[10]参见Lewis的"幸存和同一性"以及我的"刘易斯、佩里和什么重要"，两者皆见于Rorty。

[11]总体上说来我所赞同的其他还原论者包括：H·P·格赖斯（见Perry（1））；A·J·艾耶尔（特别参见"人的概念"），见Ayer（1）；A·昆顿；J·L·麦基（见Mackie（4）和（5））；J·佩里，尤其是"是同一的之重要性"，见Rorty and Perry（2）；D·K·刘易斯（见Rorty）；以及S·休梅克（见其《个人同一性》，布莱克韦尔出版社，1984年。）

[12]Reid，见Perry（1），第114页。

[13]见Williams（9），重印载Williams（2），第19—25页。

[14]Shorter；以及J.M.Jack（未发表），他要求把这个标准嵌入一种因果理论。

[15]Williams（2），第20页。

[16]Wiggins（1）、（2）和（3）提出相似的一些论证。对于所引发的一些问题，此处我未予讨论，对此Nozick（3），第656—659页上有清晰的讨论。

[17]Williams（2），第23页。

[18]参见Wiggins（1），第55页。

[19]Wiggins（4），第146页。正如我们下面将看到的，一些作者拒斥这个主张。

[20]Williams（9），重印载Williams（2），第24页。

[21]见Wiggins（3）。遗憾的是，威金斯在此并未继续讨论Wiggins（1）中他所想象出的那个分裂事例。

第十三章

到底什么重要

95.从自我中解脱出来

真相与我们所倾向认为的大为不同。即使我们没有觉察到此点，但是我们绝大多数人是些非还原论者。如果我们考虑一下我所想象出的那些事例的话，我们会强烈地倾向于认为我们的连续存在是与物理的和心理的连续性不同的一个深层的进一步事实，是一个必定要么全有要么全无的事实。真相并非如此。

真相令人感到压抑吗？一些人可能发现如此。但是我却发现它是令人感到解放和感到欣慰的。当我过去认为我的存在是这样一个进一步事实的时候，我就好像囚禁在自我之中。我的生活就像一个玻璃隧道，我在其中一年比一年快地穿行着，而在隧道的尽头是一片黑暗。当我改变了我的观点的时候，我的玻璃隧道的墙壁就都消失了。我现在生活在露天之中。在我的生活和他人的生活之间仍然存在着一种差异。但是差异变得小了。他人更加近了。我对自己余生的关心比以前少了，而对他人生活的关心则比以前多了。

当我过去相信非还原论者的观点的时候，我还对我不可避免的死亡更加关心。在我死亡之后没有一个活着的人将是我。现在我能够对这个事实加以重新描述。尽管后来将有很多经验，但是这些经验中没有哪一个经验将通过诸如经验-记忆中所包含的那样的直接联系链，或者像实施一个早先的意向中所包含的那样的直接联系链与我现在的经验直接联系。这些未来的经验中的某一些经验可能通过一些较不直接的方式与我现在的经验相关联。后来将有一些关于我的生活的记忆。而且后来可能有些受到我的思想影响的思想，或者有些遵照我的建议所行带来的结果。我的死亡将打破我现在的经验与未来的经验之间的更加直接的关系，但是将不打破其他各种各样的关系。就没有一个活

着的人将是我这个事实而言，这就是全部。既然我已经看清了这一点，那么我的死亡在我看来就不再那么糟了。

代之以说“我将死亡”，我应该说“将没有什么与现在的这些经验以某些方式相关联的未来经验”。因为这个重新描述让我想起这个事实所牵扯的，它遂使这个事实变得不再那么令人压抑。下面假设我必定经受某个严酷的折磨。代之以说“受难的人将是我”，我应该说“将有与现在的这些经验以某些方式相关联的受难”。经过重新描述的事实再次在我看来不再那么糟。

通过活灵活现地想象我将要经受我所描述过的那个手术，我能够增强这些效果。我想象我在联合谱系的一个中间事例中，在这个事例中我是否将要死去的问题是一个空洞的问题。难以相信该问题会是空洞的。当我回顾支持这个信念的那些论证并且又使我自己确信的时候，这使我对我的未来的自然关心暂时中断。当我的实际未来将是相当冷酷的时候——就像我将遭受酷刑或者在拂晓面对行刑队会出现的情况那样——我具有这种方式来暂时中断我的关心将是件好事。

在殚精竭虑地思考其论证之后，休谟被抛入“人们所能够想象出的最为悲惨的境地，四周是无尽的黑暗”[1]。排遣之道是享用美食和与朋友玩15子棋戏。休谟的那些论证支持的是完全的怀疑主义。这就是为什么它们带来的是黑暗和孤寂的原因。那些支持还原论的论证对我却产生了相反的效果。经过殚精竭虑地思考这些论证，我与他人之间的玻璃墙消除了。而且正如我已经说过的，我对我的死亡较少关心了。这只不过是这样一个事实，即在某段时间之后将发生的经验中没有哪一个经验将以某些方式与我现在的经验相关联。这能那么重要吗？

96.躯体的连续性

因为它以这些方式影响到我的情绪，还原论者的观点为真这一点令我欣喜。这只是一个心理效果的报告。发生在其他人身上的效果可

能有所不同。

还有其他几个有待讨论的问题。而且在讨论这些问题的时候我所能够做的远远不止是报告那些关于我的反应的事实。对这些问题的回答部分依赖于某些论证的力量。我将首先讨论，作为还原论者我们应该主张什么是重要的。然后我将询问，如果我们已经改变了有关个人同一性的性质的观点，我们应该如何改变我们有关合理性和道德性的信念。

就像我的分裂事例所表明的，个人同一性无关紧要。情形只是在绝大多数情况下个人同一性与重要的东西相重合而已。在个人同一性被误认为重要的那种方式中重要的是什么？在我们对自己的未来的关心中予以合理关心的是什么？

这个问题可以予以重述。为简单性起见，假定一个人只关心自己的自利会是合理的。假设我是一个自我主义者，而且我会以数种方式之一与某个作为结果的人相关联。会证明对这个作为结果的人的自我主义式的关心为正当的那个关系是什么？如果这个人的余生将是相当值得一过的，那么我应该想要以什么方式与这个人相关联？如果他的余生糟得无以复加，那么我应该想要以什么方式不与这个人相关联？简言之，对一个自我主义者而言从根本上说应该重要的是什么关系？对我们所有的人而言，这个关系也将是在我们对自己未来的关心中从根本上说应该重要的东西。但是既然我们可能关心这个作为结果的人的命运，他与我们的关系无论是什么，最为清楚明白的就是询问对于一个自我主义者而言重要的应该是什么。

这里是些最为简单的答案：

（1） 物理连续性，
（2） 带有正常原因的关系R，
（3） 带有任何可靠原因的R，
（4） 带有任何原因的R。

R是带有确当种类原因的心理联系性和/或者连续性。如果我们决定R是重要的，那么我们必须考虑联系性和连续性的相对重要性。或许有人提出，重要的既是R又是物理连续性。但是这与答案（2）是相同的，因为物理连续性是R的正常原因的组成部分。

我们能够辩护答案（1）——只有物理连续性重要——这一主张吗？我们能够主张如果我将与某个作为结果的人在物理上是连续的，这就是所重要的，即便我与这个人将不处于关系R中吗？

重新考虑一下威廉斯的例子，其中外科医生完全毁掉了任何可辨种类的心理连续性。假设这个外科医生准备在我身上以一种无痛的方式动手术，而且这个作为结果的人将有一种糟得无以复加的生活。如果我是一个自我主义者，我可能把这个前景当作并不比无痛死亡更糟，因为我并不关心发生在这个作为结果的人身上的将是什么。我也或许转而把这个前景当作比死亡糟得多，因为我以自我为中心地关心这个人的骇人听闻的未来。我的态度应该是哪一个？

我应该以自我为中心地关心这个人的未来，如果我能正当证明地相信这个人将是我而不是只与我在物理上连续的某个别人的话。但是正如我已经论证的，这个信念是得不到正当证明的。威廉斯的例子处于心理谱系的远端。在这个谱系的那些中间事例中和在其远端，在这个作为结果的人是我和他是某个别人之间都不存在真正的差别。这些不是其中的一个必定为真的两个不同的可能性。在威廉斯的例子中，全部事实就是这些。那个作为结果的人将从物理上与我相连续，但是并不从心理上与我相连续。我们可以把这个人称作我，也可以称之为某个别人。照同一性的广义心理准则，我们可以称他为某个别人。但是这两个描述中没有哪一个会是一个事实性的错误。两者都是对同一事实的描述。如果我们为了蕴含某个有关什么重要的观点而给出这些描述之一，那么我们的描述或许就是一个坏的描述。它或许蕴含一个有关什么重要的不可辩护的观点。但是我们必须在选择我们的描述之前决定什么是重要的。

假设我接受这些主张。作为一个还原论者，我应该以自我为中心地关心这个人的未来吗？尽管我知道物理连续性并不能使这个人将是我这一点为真，但是我应该将之作为一个进一步事实关心吗？在决定什么重要的时候，我必须搁置有关我的同一性的所有观点。有关同一性的问题在此是空洞的。我必须问物理连续性本身是否证明自我中心主义的关心是正当的。

我认为答案应该是否。正如我所论证过的，那些相信物理准则的人们不能令人信服地要求整个躯体的连续性。如果某个移植器官发挥的功能一样好，我是否移植了某个器官这一点不可能是重要的。能被主张为重要的全部东西是我的足够大脑的连续存在。

为什么应该以这种方式单单挑出大脑？答案必定是：“因为大脑是心理连续性或者关系R的载体。”如果这是为什么大脑被单单挑出来的原因的话，那么大脑的连续性当不是关系R的载体的时候就不会是重要的。大脑的连续性在此处绝不会比躯体上的任何其他部分的连续性更加重要。而这些其他部分的连续性毫不重要。如果这些其他部分被充分相似的复制品所置换，不会重要。关于大脑我们应该持同一主张。如果大脑的连续性将是关系R成立的原因，那么它重要。如果R将不成立，那么大脑的连续性对于其大脑是现在所谈论的大脑的那个人而言应该没有什么重要意义。它不会为自我主义式的关心提供正当证明。

还原论者们不能令人信服地主张惟有物理连续性重要。他们充其量能够主张这个连续性是重要的东西的一部分。他们充其量能够辩护答案（2），即如果关系R不具有其正常成因——物理连续性是其中的一部分——的话，R就不会有什么重要了。

我认为（2）也是不可辩护的。我认为物理连续性是个人连续存在中最不重要的因素。我们在我们自己和他人中所珍视的不是同一特定大脑和躯体的连续存在。我们所珍视的是我们自己与他人之间的各

种各样的关系，是我们所爱的人和物，是我们的雄心、成就、承诺、情绪和记忆，以及其他一些心理特征。我们中的某些人还想要我们自己或者他人连续具有与我们现在的躯体非常相似的躯体。但这不是想要同一特定的躯体连续存在。我相信，如果以后将有某个人与现在的我处于关系R中，那么这个人是否具有我现在的大脑和躯体就几乎一点也不重要。我认为从根本上说重要的是关系R，即便它并不具有其正常原因。于是，如果我的大脑被一个精确的复制品所置换，也不会重要。[2]

如果某个人将与我处于R关系中，这个人的躯体也应该充分地像我现在的躯体，以允许完全的心理联系性。这不会为真；例如，如果这个躯体是异性的躯体。而且对一些人而言，诸如对那些非常漂亮的人中的某些人而言，也应该有确切的物理相似性。关于该相似性的这些主张今后我将予以省略。

我们是否接受这个观点可能影响到我们关于自己的生活的信念和态度。但是该问题在电子传输这个想象出的事例中是最为清楚明白的。照我的观点，我与我的复制品之间的关系包含从根本上说重要的东西。这个关系大约像通常的幸存一样好。依我之见，从非还原论者的观点之立场来判断，通常的幸存比被摧毁和被复制好不到哪里去，甚或如同被摧毁和被复制一样糟。因此，为一趟传统的太空飞船之旅付出高得多的费用会是非理性的。

很多人会害怕电子传输。我承认，在某种程度上我可能害怕。但是正如我所论证过的，这样的害怕不可能是合理的。既然我确切地知道将发生的是什么，我不可能担心两个结果中较糟的那一个将是会发生的这样一种情况。

我与我的复制品之间的关系是不带有其正常原因的R关系。原因的这种非正常性在我看来是微不足道的。考虑一下会恢复失明者视力的人工眼睛。假设这些眼睛会赋予这些人视觉，就像正常视力中所包含的那些视觉一样，那么这些视觉会提供有关能够看到什么的真实信

念。这必定会与正常的视力一样好。因为这些眼睛不是人的视力的正常原因而拒斥这些眼睛会是不可取的。不喜欢人工眼睛的理由会不少，因为它们会使一个人的外表让别人感到不安。但是在电子传输中却没有这样的类比性。我的复制品尽管是人工产物，但是他将在每一个方面都像我。他将具有正常的大脑和躯体。

我们很可能决定不称我的复制品为我。如果我们作出这样的决定，我们应该把这个事例当作像我的分裂一样。这是对一个空洞问题有一个最佳答案的一个事例。如果我将要分裂，最好说作为结果的两个人中没有一个是我。但是这并不蕴含着我应该把分裂当作死亡。我们不是在决定数个结果中哪一个将是会发生的；我们只是在从对一个单一结果的数个描述中选择一个描述而已。既然如此，那么我们对一个描述的选择与我如何看待这个结果这一问题并不相关。

在那个我将被电子传输的事例中情况也是如此。我对这个结果的态度不应该受到我们是否称我的复制品为我这一决定的影响。即便我们尚未作出这个决定，但是我知道全部事实。如果我们的确决定不称我的复制品为我，那么事实是：

（a） 我的复制品将不是我

就会寓于事实

（b） 将没有物理连续性

和事实

（c） 因为如此，R将没有其正常原因。

既然（a）会寓于（b）和（c）之中，我就应该忽略（a）。我的态度应该取决于事实（b）和（c）。对我的复制品不是我而言，这些

事实就是全部。

当我们看清最后这个主张为真的时候，我相信我们不能合理地主张（c）相当重要。原因非正常这一点无关紧要。重要的是效果。而这个效果——关系R的成立——本身是相同的。诚然，如果这个效果具有正常原因，我们能够以一种不同的方式予以表述；我们能够说，虽然我的复制品与我在心理上相连续，他却不是我。但是，这并非是在所发生的东西中超出原因差异之外的一个进一步差异。如果我决定不按下那个绿色按钮，而是宁愿为一趟传统的太空飞船之旅付出多得多的费用，那么我必须承认这只是因为我并不喜欢想到一个因果关系的一个非正常途径。对此原因的非正常的大为关心不可能是合理的。

同样的评述也适用于一个人现在的大脑和躯体的连续存在。想要使我的复制品的躯体像我目前的躯体，可能是合理的。但这是一个对某类躯体的渴望，不是一个对同一特定躯体的渴望。我们为什么应该想要这个大脑和躯体真正抵达火星？而且，这种自然而然的担心在于只是确保我将抵达火星。但是这再次假定我是否抵达火星的问题在此是一个真实的问题。而我们应该断定这是一个无意义的问题。即使这个问题具有一个最佳答案，我们还是能够在决定这个答案是什么之前就确切地知道所发生的将是什么。既然如此，我能合理地大为关心这个人的大脑和躯体是否将是我目前的大脑和躯体吗？我认为，虽然稍加关心可能并非是非理性的，大为关心则会是非理性的。

稍加关心为什么不会是非理性的？这就像一个人想保留同一个婚戒，而不是一个与之完全相同的新戒指一样。我们理解出于感情保留婚礼上的那个戒指的愿望。同样，适度偏爱火星上的那个人具有我目前的大脑和躯体可能并非是不合理的。

还剩下一个问题。如果将有一个与我处于关系R中的人，那么如果这个关系并不具有一个可靠的原因这一点重要吗？

对原因将是可靠的这一点事先表现出偏爱是有明显理由的。假设电子传输在少数情况下完美地运作过，但是在绝大多数情况下却是彻底地失败。在少数情况下，火星上的那个人会是我的一个完美的复制品。但是在绝大多数情况下，他会完全不像我。如果这些是事实的话，为太空飞船旅行付出更多的费用就明显地是合理的了。但这是毫不相关的。我们应该问:“在为数不多的那些我的复制品与我处于关系R中的事例中，R如果没有一个可靠原因的话会重要吗？”

我认为答案再次应该是否。假设对某种疾病有一个不可靠的治疗方案。在绝大多数情况下，这个治疗方案毫无建树。但是在一些情况下，它却把疾病完全治愈。在这些情况下，只有效果重要。其原因纵然是不可靠的，但是这个效果一样好。关于关系R我们应该持同一主张。我断定，就我所表述的那些答案而言，我们应该接受答案（4）。在我们对自己未来的关心中，从根本上说重要的是不管哪种原因导致的关系R。

97.支线事例

我已经论证过，电子传输几乎会像通常的幸存一样好。对这一主张的另一个挑战来自支线事例。假设新扫描仪没有毁坏我的大脑和躯体，但是损害了我的心脏。我在地球的此地，预计不日会死去。通过使用内部通话设备，我看到火星上我的复制品并且与之进行了交谈。他向我确保将从我生命停止的地方继续我的生命。

我此处的态度应该是什么？我要死了。我与这个复制品之间的关系包含重要的东西吗？他不是与现在所是的我，而是与今天早晨按下那个绿色按钮时的我在心理上完全相连续。这个关系几乎像幸存一样好吗？

人们可能难以相信是这样。但是人们也难以相信我的生命是否短暂地与我的复制品的生命相重叠也可能相当重要。

考虑一下下面的事例可能有所帮助:

安眠药。实际上，某些安眠药可能引发退化性健忘症。情况是，如果我服用这样的一粒药，在长达1个小时的时间内我将仍然保持醒觉，但是在夜眠之后我将没有对这1个小时的后半个小时的记忆。

事实上我已经服用了一些这样的药，而且发现了结果怎样。假设我在接近1个小时之前服用了一粒这样的药。明天在我的床上醒来的那个人将不与现在所是的我在心理上相连续。他将与半个小时前所是的我在心理上相连续。现在我是在一条心理支线上，它很快将在我入眠时终结。在这半个小时期间，我与我过去的自己在心理上相连续。但是我现在并不与未来的我自己在心理上相连续。随后我将永远不记得在这半个小时期间的所行、所想和所感。这意味着在某些方面我与明天的我自己的关系就像与另一个人的关系一样。

比如，假设我一直对某个实际问题忧心忡忡。我现在看到了解决方案。因为我应该做什么是清楚明白的，所以我形成了一个坚定的意向。倾我之余生，足够形成这个意向。但是，当我处在这个心理支线上的时候，这是不够的。后来我将不记得我现在已经决定了什么，而且我将不带着我现在已经形成的意向醒来。因而我与明天的我自己的联络必定如同与某个别人的联络。我必须给我自己写一封信，表述我的决定和新的意向。然后我必须把这封信放到我明天铁定会注意到的地方。

事实上我并不具有对于作出这样的一个决定和写下这样的一封信的记忆。但是我的确曾经在我的剃须刀下面发现了这样的一封信。

这个事例在某个方面像那个支线事例。那个事例就像这个事例的一个变体。在那个事例中，我在离开那个舱室之后活了一些时日，在我死去之前我的复制品不会被创造出来。但是在我们正在考虑的这个事例中，我的生命与我的复制品的生命相重叠。我们在内部通话设备上交谈。在安眠药事例中，没有这样的类比性。

在那个我想象出的物理考试事例中，能够发现这样的类比性。在该事例中，我的头脑分裂了10分钟。在我的两个意识流中，我都知道我现在具有我的另一个意识流中的思想和感觉。但是在各自意识流中，我没有觉察到在我的另一个意识流中的我的思想和感觉。我与在我的另一个意识流中的我自己的关系再次像我与另一个人的关系。我会不得不以一种公共的方式进行联络。我或许在一个意识流中给在我的另一个意识流中的我自己写一封信。然后我会用一只手把这封信放到另一只手中。

这就像我在支线事例中的境遇一样。我能够想象具有一个分裂的头脑。既然如此，我无需假定在火星上的我的复制品是某个别人。在

这里的地球上，我现在并不知道在火星上的我的复制品正在想什么。这就像这样一个事实：在那个物理考试事例里，在我的两个意识流的每一个中，我并不知道在我的另一个意识流中我在想什么一样。我能够相信我现在的确具有另一个意识流，对此现在我在这个意识流中并不知道。而且，如果这有所帮助的话，关于我的复制品我能够秉持这样的观点。我能够说，我现在具有两个意识流，一个在这里的地球上，另一个在火星上。这个表述不可能是一个事实性的错误。当我与火星上我的复制品交谈的时候，这只不过像物理考试事例中在我的两个意识流中的我自己之间的交流。

这一安眠药实例提供了与支线事例的特殊特征之一的一个切近类比，即我处于一个心理支线上。那个想象出的物理考试事例提供了与另一个特殊特征的一个切近类比，即我的生命与我的复制品的生命相重叠。当我们考虑这些类比性的时候，这看起来足以辩护这样一个主张，即当我处于支线上的时候，我与我的复制品之间的关系几乎包含一切重要的东西。有一点可能稍微有点麻烦，这就是我的复制品将与今晨我按下那个绿色按钮时的我在心理上相连续，而不是与作为现在的我在心理上相连续。但是，这些关系本质上是一样的。些微差异是我的生命与我的复制品的生命略有重叠。

如果重叠巨大的话，这就会造成不同了。假设我是一位将要谢世的老人。某个曾经是我的复制品的人将活得比我长。当这个人40年前开始存在的时候，他在心理上与那时的我相连续。自那以后他已经独自生活了40年。我同意我与这个复制品的关系固然比通常的死亡要好一些，却并不是几乎像通常的幸存一样好。但是如果我的复制品与10天前的我或者10分钟之前的我在心理上相连续的话，这个关系就会几乎一样好了。正如诺齐克所论证的，像这样短暂的重叠不能被合理地认为具有相当重要的意义。[3]

尽管我的两个类比看似足以辩护这个主张，但是我坦白承认这是那些我的观点最难以令人相信的事例之一。在我按下那个绿色按钮之

前，我更容易相信我与我的复制品的关系包含着在通常的幸存中根本上重要的东西。我能够沿着主线进行前瞻，有40年的寿命在前头。在我按下那个绿色按钮并且与我的复制品交谈之后，我不再能够以同一方式沿着主线进行前瞻。我对未来的关心需要改变方向。我必须努力把这个关心一直回溯过支线上的那个分裂点，然后沿着主线前瞻。心理上的这个机动会是困难的。但并不是令人惊讶的。而且既然不是令人惊讶的，这个困难就不提供一个针对我就这个事例所主张的充分论证。

98.系列-个人

我已经否定了个人同一性是重要的。依我之见，在我们对未来的关心中根本上重要的是不管哪种原因导致的关系R的成立。即使当关系R并不与个人同一性重合的时候，这也会是重要的。

照内格尔的观点，我所是的在实质上是我的大脑。而且根本上重要的是这个大脑的连续存在。我认为在"附录"D中我回答了内格尔的观点。但是这些论证可能并不令人信服。因此，值得解释何以内格尔的观点和我的观点的一个修正形式两者都会为真。

假定内格尔的观点为真。对我来说，根本上重要的是我的大脑的连续存在。既然照内格尔的观点我实质上是我的大脑，那么我就不可能决定采取关于我自己的一种不同的观点。但是我能够做点别的。

内格尔表述了系列-个人这么一个概念。照内格尔的观点，一个个人实质上是一个特定的具体化的大脑，而一个系列-个人潜在地是一些具体化的大脑的一个R-相关系列。我现在不能够解决我们的躯体老化和腐朽的难题。内格尔想象出这样一个共同体，其中技术提供了一个解决方案。在这个共同体中，每一个超过30岁的人都要一年一度地进入一个扫描复制器中。这个机器毁掉人的大脑和躯体，生产出一个复制品，这个复制品与这个人R-相关，并且具有一个除没有老朽之外都确切相似的躯体。内格尔主张，对这个共同体中的那些系列-个人而

言，使用这样的扫描复制器不是非理性的。照他们的同一性准则，这些系列-个人中的每一个都会继续存在，每年都移至一个新大脑和新躯体。每一个系列-个人都总会具有一个大脑和躯体，这个大脑和躯体具有他30岁时大脑和躯体的青春、外貌和活力。

这些系列-个人或许会遭遇致命的意外。因此，我添加这样一个细节，即作为一种防范措施，每一个人每天都使自己的图像被复制下来。有了这个添加，这些系列-个人都是潜在地不朽的，而且会有永恒的青春。照现在被接受的绝大多数理论，宇宙要么将无限地扩张，要么将在大爆炸的逆转中坍缩。绝大多数物理学家假定，无论照两者中的哪一个情形，所有的生命形式将变得不可能。如果这情形并不为真的话，那么这些系列-个人就会永生了。[4]

假定我接受内格尔的观点。我实质上是我的大脑，在我对我的未来的关心中根本上重要的是这个大脑的连续存在。如果我接受这个观点，那么必定是不情愿地被迫得出这个结论。我相当偏好我的老观点。尽管我不能够改变我关于我是什么的观点，但是我能够做下述的事情。你现在正在阅读我1982年11月写下的句子。这个句子告诉你在本书的所余部分，代词被用来指代系列-个人。如果内格尔的观点是虚妄的，这可能不会改变代词所意指的。每一个人都可能是一个系列-个人。如果我们的同一性准则是不管哪种原因导致的关系R的话，就会如此；因为这也是系列-个人的同一性准则。

如果内格尔的观点为真，那么本书所余部分是在一个新的意思上使用代词。它们指代的不是德里克·帕菲特这个个人，而是系列-个人。这样，“主格我”（I，主词）和“宾格我”（me，宾词）并不指代德里克·帕菲特这个个人。他们指代系列-个人，其目前的大脑和躯体也是德里克·帕菲特的大脑和躯体。既然单词“主格我”和“宾格我”是在这一新意思上使用的，那么它们的旧有意思是以单词“旧主格我”和“旧宾格我”来表达的。其他代词同理亦然。

照内格尔的观点，旧宾格我和系列-个人的我之间的关系是什么？记起一个神话中的生物——凤凰——可能有所帮助。照鸟的同一性准则，如果一只鸟被烧成灰烬它就不复存在。如果凤凰存在的话，它就不会是一只特定的鸟。它会是一系列的鸟，或者说一只系列-鸟。凤凰在任何时候都有某个特定的鸟的躯体。但是当这个鸟被烧成灰烬的时候，只是这个特定的鸟不复存在。凤凰从灰烬中升起，在一个新的鸟体中获得重生。像一个特定的系列-个人一样，一个特定的凤凰于是会有一系列不同的躯体。

凤凰从未这样存在过。但是有很多系列-个人。这些句子正在被一个系列-个人——宾格我——写下来。它们也正在被一个个人——旧宾格我——写下来。这个个人被命名为德里克·帕菲特。作为系列-个人我据此把我自己命名为凤凰·帕菲特。既然我目前的躯体也是德里克·帕菲特的躯体，我们两者都正在写下这些句子。而且我们两者都具有同样的思想和经验。但是，尽管我们是以这种极端紧密的方式联系在一起，如果内格尔的观点为真的话，我们是两个不同的个体。我们之间的区别是这样的。照内格尔的观点，如果旧主格我被电子传输的话，这就会杀死旧宾格我，即作为那个个人的我。但是它不会杀死宾格我，即作为系列-个人的我。这个差异足以使旧宾格我与宾格我成为两个不同的个体。

你会怀疑我即这个系列-个人是否真的存在。你会认为仅仅凭发明一个新概念不能导致事物的存在。的确如此。我既不是由内格尔发明的系列-个人概念而导致存在的，也不是由写下上述那些可能已经赋予“主格我”和“宾格我”以新义的句子而导致存在的。鉴于用“系列-个人”概念所指的意思，当作为那个个人的旧主格我存在的时候，作为系列-个人的主格我开始存在。而且很有可能的是两者将同时不复存在。这之所以是很有可能的，是因为电子传输在旧主格我的有生之年将会成为最没有可能的。

语言是发展的。当我们发明一个新概念的时候，我们可能发现它适用于部分实在。凤凰这个概念不适用于任何东西。但是系列-个人这个概念就像个人概念一样经常地适用。无论内格尔的观点是否真实，现在存在着数十亿系列-个人。如果内格尔的观点真实，那么对每一个存在的个人而言都有一个与这个个人非常紧密地相关的系列-个人存在。鉴于这一极端紧密的关系，这些个体之间几乎从不值得加以区别。只有在我讨论什么重要这一点上才值得加以区分。我在假设内格尔的观点为真。我在假设对旧宾格我而言根本上重要的是我目前的大脑的连续存在。即使我相信这一点，我仍然可以认为我们目前大脑的连续存在不是重要的。我们能够主张重要的是关系R。这是对我们即诸系列-个人来说重要的。

借助某个新概念，我们有时候能够对实在进行更好的描述。原子和分子概念就是如此。这类改进是直截了当的。如果内格尔的观点为真，系列-个人概念也使对实在进行更好的描述成为可能。但是这种改进就并不是这么直截了当了。系列-个人概念不仅能够使同一些人更好地描述实在，而且能够使不同的一些人既宣告他们的存在又作出这样的更好描述。

内格尔提到了第三个概念，即白昼-个人概念。这样的一个人的存在必然涉及一个不间断的意识流。对白昼-个人而言，睡眠等于死亡。当我们考虑这个概念的时候再次发现它是适用于实在的。在任何时候有多少有意识的人活着就有多少白昼-个人活着。但是在整个一年中，生活过的白昼-个人的数量将远远超出活着的个人的数量。

白昼-个人概念比个人概念要糟一些。这是因为重要的不是意识流的不间断性。所重要的是关系R。尽管白昼-个人这个概念适用于实在，但是它以并不重要的界线区分出部分实在。我们令人信服地相信，在意识流中如果有间断也不重要。这之所以并不重要，是因为这些间断并不毁掉心理的连续性。

如果内格尔的观点为真，对于一个个人而言重要的是一个特定大脑的连续存在。但是如果我问关于我们自己、我们的生命和我们彼此之间的关系重要的是什么，特定大脑的连续存在看起来就并不是重要的。毋宁说，正如我已经主张的，更加重要的是关系R，是带有无论何种原因的心理的联系性和/或者连续性。如果是这样，而且内格尔的观点为真，那么个人概念比系列-个人概念要糟，而且是在类似的意义上要更糟。照内格尔的观点，对一个个人而言所重要的是他的目前大脑的连续存在。对一个系列-个人而言所重要的是带有无论何种原因的关系R。系列-个人概念要好一些，因为它适用于更加重要的东西。

一个个人不能否认他是一个个人。而且，如果内格尔的观点为真，一个人不能把自己转化为一个系列-个人。但是一个系列-个人可以通过两者共享的嘴巴说话。这个系列-个人能够宣告他的存在，给自己命名，而且声称他写出或者说出的所有代词指代系列-个人。所有的其他系列-个人也能如此。所有的未来人的生活然后会被那些把他们自己当作系列-个人的人们所过着。这些生活也会被那些个人所过着。但是诸个人现在会有一个从属的地位，因为他们几乎不会指称旧的他们自己。

如果内格尔的观点不为真，我刚刚描述的事件可能不会造成任何不同。我们关于同一性的准则的信念可能没有涵盖一些实例，诸如那些具有分裂了的半脑的人们等等。而且这些信念显然没有涵盖许多想象事例。既然人们不是与他们的大脑、躯体和经验相分离的独立存在的实体，那么在这些想象事例中有关个人同一性的问题就是空洞的。在这些事例中我们能够赋予这些问题以答案，据此展开我们有关个人同一性准则的信念。一个可能的展开会使这个准则成为不管哪种原因导致的关系R的非分支形式的成立。照这个准则，个人都是些系列-个人。上述所作的区别消失了。

如果内格尔的观点为真，我不能持这些主张。照他的观点，每一个个人实质上都是他的大脑，而且对每一个人而言根本上重要的是这个大脑的连续存在。如果是这样，个人并非系列-个人。我上面描述的事件就会造成区别。如果所有的系列-个人都宣告他们的存在，而且开始使用代词来指代他们自己的话，这就会是一种改进。每一个系列-个人都是与某一个特定的个人非常紧密地相关的。如果在每一个这样的对子中系列-个人占据主导角色，就会好得多。系列-个人概念以一种更少随意性的方式区分出实在的那些部分。作为系列-个人，我们能够否定对我们而言所重要的是我们的大脑的连续存在。我们能够主张，正如我已经论证过的，根本上重要的是不管哪种原因导致的关系R。

99.我是一个迹象还是一个类型?

威廉斯考虑了一个其中一个个人会有许多共存的复制品的事例。而且他提出了对这类事例的一个新描述。他描述了一个个人-类型概念。假设有某个特定的个人玛丽·史密斯。而且假设扫描复制器生产出某个特定时刻的玛丽·史密斯的许多复制品。这些复制品将都是玛丽·史密斯。她们将是同一个个人-类型的不同迹象或例证。如果这样的情况发生的话，就会有多个有关什么重要的问题。假定该事例涉及那个毁掉了玛丽·史密斯的大脑和躯体的旧扫描仪。在她按下那个绿色按钮之前，有关她与她未来的诸复制品之间的关系，玛丽·史密斯应该相信什么？所重要的是她目前的大脑的连续存在吗？如果后来将出现她的类型的一些活生生的迹象，那么这与她目前的大脑连续存在这种情况大致一样地好吗?

这是一个对于原初的玛丽·史密斯而言什么应该重要的问题。威廉斯并未讨论这个问题。但是他有趣地谈到另一个问题：

> 既然我并不是在假设迹象-个人一经从原型拓印下来就具有相互交流的经验……他们将受到不同经验的各种影响，而且将倾向于日益相异。被视作原型的复本，他们将变成日益模糊和被改写的复本；凭其本身看来，他们将日益变成一些个体人格。这或许是受欢迎的。因为

某个爱过这些迹象-个人之一的人，爱她有可能不是因为她是某个玛丽·史密斯，而是不顾她是某个玛丽·史密斯这个事实……。玛丽·史密斯分化得越多，有情人对于他情有独钟的东西可能就感到越加有把握，就越放心可靠。

如果有人爱恰像某个玛丽·史密斯的一个迹象-个人，那么该迹象-个人是否是他的真正所爱这一点有可能并不清楚。他所爱的是玛丽·史密斯，就是说爱的是那个类型-个人。我们能够隐隐约约地看到这会像什么。这就会像爱一件质地为可再生的艺术品。一个人可能开始比较该类型的，不妨说，各种表演；想要接近某个所钟情的人就会像非常想要聆听某场《费加罗的婚礼》的演出一样，即便那是一场平常的演出——恰如一个人将去聆听地方剧团勉强凑合的《费加罗的婚礼》的演出，而不是根本不听《费加罗的婚礼》一样，一个人宁愿看到当地打了折扣的玛丽·史密斯，而不是根本不看玛丽·史密斯。

借此，我们所称的爱一个人这一点本身就开始出现裂解，而且对这一点的反思可能鼓励我们不去贬低我们实际上所处的那个深深地以躯体为基础的状况。固然，在目前的情形下爱一个人并不确切地像爱一个躯体一样，说它们基本上是同一回事比犯一个深刻的形而上学的错误大概更加具有荒诞的误导性；但是，如果它听起来并不很高尚的话，从悬置我们目前的情形中而如此兴旺地生长出来的那些选择听起来也并不怎么精神味实足。[5]

爱一个人与爱那个人的躯体基本上就是同一回事吗？威廉斯承认，这个主张是"具有荒诞的误导性的"；但是他提出，它比任何一个别的选择都要好。而且他警示我们不要贬低"我们实际上所处的那个深深地以躯体为基础的状况"。与之不同的一个其中个人会有许多共存的复制品的情形，会威胁到相当多的我们所珍视的东西。

我认为我们应该接受最后这个主张。但是我们不应该主张爱一个人与爱那个人的躯体基本上就像一回事。存在一个更好的选择。

威廉斯所作的推理可能如下。除非我所爱的是某个特定的躯体，否则我就不能够爱一个个体。假设我爱那个原初的玛丽·史密斯。一部机器毁掉了她的大脑和躯体，而且生产出一个复制品。如果我对玛丽·史密斯的爱转到她的复制品上的话，这就提示我所爱的不是一个个体，而是一个个人-类型。而且如果我们考虑这样的爱所涉及到的是什么的话，我们的发现是令人担忧的。

我同意这样的爱会是非常不同的，而且是令人担忧的。但是我拒斥刚才给出的推理。

我们应该考虑两种想象事例。其中之一是内格尔所想象出的那个共同体。在这个共同体中，尽管有复制，但是关系R从未呈现为一种分支形式。在她30岁以后，玛丽·史密斯像其他许多人一样一年一度地使用那个留住青春的复制品。如果这样的一些机器存在的话，生产出一个单一个体的数个共存的复制品会是可能的。但是我们可以假设，在这个想象出的共同体中，那些个体决定不使这样的可能性发生。由于我在第90节中所表述的各种理由，以及威廉斯更好地表达的各种理由，这些个体相信，分裂并不像通常的幸存一样好。

假设我是一个搬迁到这个与众不同的共同体的人。我爱上了玛丽·史密斯。我在她首次使用那个复制品之后应该如何反应？我主张我既会又应该爱她的复制品。这不是道德性的"应该"。照最好种类的爱的最好概念，我应该爱这个个体。她与我所爱的玛丽·史密斯在心理上是完全相连续的，而且她有一个确切相似的躯体。如果我不爱玛丽·史密斯的复制品的话，这只会是出于数个糟糕的理由之一。

其中的一个理由或许是我相信非还原论者的观点。我相信个人同一性是一个深刻的进一步事实，而这是不会由复制所产生的。我并不爱玛丽·史密斯的复制品，因为我相信在这种深刻的方式上她不是玛丽·史密斯。这个反应得不到正当证明，因为没有这样的进一步事实。

下面假设我接受还原论者的观点。但是我认为复制几乎像通常的死亡一样糟。当玛丽·史密斯按下那个绿色按钮的时候，我的反应是悲伤的。也许后来我会渐渐爱上玛丽·史密斯的复制品。但是鉴于我对发生在玛丽·史密斯身上的糟糕的事情的信念，我的爱不可能是毫无悲伤的简单转移。

正如我已经论证的，这个反应也得不到正当证明。照还原论者的观点，我们应该把复制当作几乎与通常的幸存一样好。既然玛丽·史密斯选择了被复制，我们能够假定这也是她的观点。照此观点，我的爱应该转移到她的复制品身上。况且，在这个想象出的共同体中，那些

个体是些系列-个人。既然如此，玛丽·史密斯的复制品就是玛丽·史密斯。

如果我的爱不转移，就会有两个进一步的解释。威廉斯提出，爱一个人基本上是爱一个特定的躯体。但是这种爱或者肉欲充其量是极端罕见的。更加常见的是一种对某个个人的躯体的纯粹身体上的或者性的着魔，是一种并不关涉这个人的心理的着魔。但这不是对一个特定躯体的爱。正如昆顿所写，在这样的着魔事例中，"不需要任何特定的躯体，只有一个划界精确度或多或少的种类"[6]。假设我在身体上对玛丽·史密斯的躯体着魔。这种着魔会转移到玛丽·史密斯的复制品身上。这就会像这样一种情况，其中我所着魔的躯体是一个同卵双胞胎。如果这个双胞胎死掉的话，我的着魔会转移到另一个双胞胎的躯体上。

通常的爱不会如此转移。这样的爱关涉所爱的人的心理，以及这个个人持续变化着的精神生活。而且爱一个人是个过程，不是一个固定的状态。彼此相爱涉及一段共享的历史。这就是为什么，如果我已经爱上玛丽·史密斯数月或数年的话，她的位置不会简单地被她的同卵双胞胎中的另一个所取代的原因。她的复制品的情况则大为不同。如果我已经爱上玛丽·史密斯数月或数年的话，她的复制品将具有我们所共享的那段历史的全部类记忆。

我已经提出过这样的主张，即如果我并不爱玛丽·史密斯的复制品的话，那么对此的解释不大会是我爱的是她的特定躯体。有人会有这样的爱或者肉欲这一点是值得怀疑的。剩下的解释是我的爱已经毫无理由地不复存在。没有理由是个坏理由。爱能够像这样不复存在，但这只能是一种劣等的爱。

我已经讨论了内格尔想象出来的用以替代实际世界的选择。在这个替代选择中，人们时常被复制，但是从未有某个人的两个共存复制品。关系R永远不以分支的形式呈现。我主张，在这样的世界中对一

个人的爱应该直接转移到这个个人的复制品上。这种爱之所以应该转移，是因为这个人与他的复制品的关系包含着在通常的幸存中极其重要的东西。

威廉斯提示，在一个带有复制的世界中，我们应该区别个人-类型与这些类型的迹象。但是在刚刚描述的那个世界中，关系R总是以一对一的形式出现，这就不会是一个有效的区分。如果我们把每一个新的复制品称作某个个人-类型的另一个迹象的话，我们就会错误地描述所发生的东西。这种描述会忽略至关重要的东西，即心理的连续性和生命的发展。我们会以两个方式中的任何一个方式对所发生的进行更好的描述。

如果内格尔的观点是虚假的，我们的个人同一性准则会展开为使关系R的非分支形式的成立。那么内格尔所想象出的共同体中的每一个个体就会是一个个人。这些人会一年一度地迁居到新的躯体中。既然玛丽·史密斯会有一些这样的新躯体，我对她的爱应该是依次直接转移的。虽然这些人会时常转换躯体，但对这些特定人们的爱不会受到威胁。

如果内格尔的观点为真，那么这个共同体中的诸个体并非是个人。他们是，而且他们相信他们自己是，系列-个人。系列-个人玛丽·史密斯每年都移居到一个新躯体中。就像以前，不存在什么对我们所珍视的那类爱的威胁。如果我爱上一个系列-个人，我不是在爱一个个人-类型。我是在爱一个特定的个体，她有一个持续不断的历程。

下面考虑用以替代实际世界的另一个选择：威廉斯想象出的那个选择。在这个世界中，某个单一个人有许多共存的复制品。威廉斯所提议的区分在此是有效的。考虑一下有50个芳龄30岁的格丽泰·嘉宝的复制品这种情况。这些会被很好地描述为一个个人-类型的不同迹象。正如威廉斯主张的，如果爱的对象是个人-类型，那么这与通常的爱大为不同。这不会是重视一个共享的历史的那类爱。

如果我生活在这样的一个世界中，而且我是其中一组复制品的话，我可能把我自己当作某个类型的一个迹象。我也许转而会把我自己当作那个类型吗？“类型”这个词的某种意义上，如果我是一个个人-类型，我就不可能会不再存在。即便现在没有我的个人-类型的一些迹象，仍然会有这个个人-类型。个人-类型甚至在宇宙毁灭之后幸存下来。这是因为，在这个意义上，一个类型是一个抽象实体，就像数字一样。我们不可能会把自己当作抽象实体。

无论按照“类型”的哪一种意思，在通常的爱和爱一个个人-类型之间都会有巨大的差异。后一种爱不能够是双向的。我可能爱某个个人-类型，但是这个个人-类型不能够爱我。一个类型不能比数字9更加能够爱。我不能够被英国美女或者新派美国女性所爱。实情或许是，对我的爱是某个个人-类型的特征之一。那么这个类型所有的那许多迹象都会爱我。

如果事实是这样的话，在我和这些迹象-个人之一之间可能存在双向的爱。双向的爱甚至存在于我与这些个人中的两个或者三个之间。但是正如威廉斯所主张的，这种爱会把我带离对个人-类型的爱。我之所以会被带离开来，是因为那些共享历史变得日益重要起来。

现在回到威廉斯的那些主要主张。他提出，尽管爱一个人就是爱一个特定的躯体这一主张具有误导性，但却有深刻的真理因素。他还提出，如果我们爱的对象不是一个特定的躯体的话，我们在爱的就会是一个个人-类型。这样的爱与通常的爱是极为不同的，而且会是令人担忧的。它会威胁到许多我们所珍视的东西。

我接受最后这个主张，但是否定其他那些主张。跟随昆顿，我对会有人爱一个特定的躯体表示怀疑。一个纯粹身体上的着魔是对一类躯体的着魔，或曰对一个躯体-类型的着魔。如此这般，它会具有爱一个个人-类型所带有的那些令人担忧的特征。一旦同卵双胞胎中的一个死亡，这种着魔会毫不悲伤地被转移到双胞胎中的另一个的躯体上。

我还否定了如果我们爱的对象不是一个特定的躯体的话我们必定爱一个个人-类型这一点。这在内格尔所想象出的用以替代实际世界的那个选择中看得最为清楚，其中人们经常被复制，但是只是以一对一的形式被复制。在这个世界中，关系R横贯许多不同的躯体，但是它从不呈现为一种分支形式。我主张，在这样的一个世界中，通常的爱会毫无改变地幸存。如果内格尔的观点为假，在这个社会中的人们尽管每年都会移居到新躯体中，但他们仍然是些特定的人。如果内格尔的观点为真，迁居到新躯体中去的就会是系列-人们。但是爱依然是对一个特定个体的爱。一个系列-个人是一个个体。

如果这些主张是正确的，我能够再次保有我所辩护的那个观点。重要的不是一个特定躯体的连续存在，而是不管哪种原因导致的关系R。

100.部分幸存

在考虑实际中的生活之前，我将首先一瞥最后一组想象事例。其一是分裂的对立面：聚合。同一性在逻辑上是一对一的，是要么全有要么全无的。恰如分裂表明，在幸存中重要的东西无需呈现为一对一的形式一样，聚合表明它能够具有不同的程度。

在联合谱系的那些中间事例中存在着聚合。在这些事例中，作为结果的那个人会在心理上既与我又与某个别人相联系，而且其程度大约相同。

我们能够想象一个其中聚合是一个自然过程的世界。两个人聚集到一起。当他们处于无意识状态的时候，他们两个的躯体长成一个躯体。然后醒来的是一个个人。

这一个个人会虚拟记得原先那两个人的生活。无需失去任何类记忆。但是某些事情必定会失去。任何两个聚合在一起的两个人都会有不同的性格特点、不同的愿望和不同的意向。这些又如何得以结合在一起?

答案或许是这样的。这些特征中的一些是相容的。它们会在那一个作为结果的人中共存。而有一些将是不相容的。如果具有同等力量的话，它们会相互抵消；如果具有不同的力量，强者将被削弱。这些效果或许就像支配显性基因和隐性基因的规律那样具有可预测性。

这里有一些例子。我钦佩帕拉弟奥[7]，而且打算访问威尼斯。我将与某个钦佩乔托[8]并且打算访问帕多瓦的人相聚合。那个作为结果的人将具有两者的情趣和意向。因为帕多瓦靠近威尼斯，两者都能容易满足。接着假设我喜欢瓦格纳[9]，而且总是投票支持社会党候选人。而另一个人痛恨瓦格纳，而且总是投票支持保守党候选人。那个作为结果的人将是一个不辨音调、立场动摇的投票人。

像分裂一样，聚合并不符合同一性的逻辑。那个作为结果的人不能被认定为与原来的两个人中的每一个都是同一个人。最佳描述是那个作为结果的人不会是这两个人中的任何一个。但是正如分裂事例所表明的，这个描述并不蕴含着这些人应该把聚合与死亡等量齐观。

他们的态度应该是什么？如果我们将要经历这种聚合的话，我们中的一些人可能把它与死亡等量齐观。而且这比把分裂等同于死亡的荒谬性要少一些。当我分裂的时候，那两个作为结果的人将确切地像我。当我聚合的时候，那个作为结果的人将并非整个相似。这使人们更加容易认定，当面对聚合的时候"我将不会幸存下来"；于是继续把幸存当作是要么全有要么全无的。

正如我已经论证的，不存在什么被涉及到的要么全有要么全无的事实。那两种联系性——物理的和心理的——会在任何程度上成立。我应该如何对待在其中这些关系的确保持在一些弱化的程度上的事例？

或许会有这样的说法："假设在我与某个作为结果的人之间会有这两个关系通常量的大约一半。这就会像通常存在的大约一半一样好。如果有这些通常量的十之八九，那么这就会有通常存在的大约十之八九那么好。"

这个观点过于粗陋。在判断某个特定聚合事例对于我具有什么价值的时候，我们必须知道我与那个作为结果的人之间的关系有多密切。我们还必须知道这个人是否将具有一些被我视作好或者坏的特征。刚才所述的观点错误地忽视了这第二个问题。

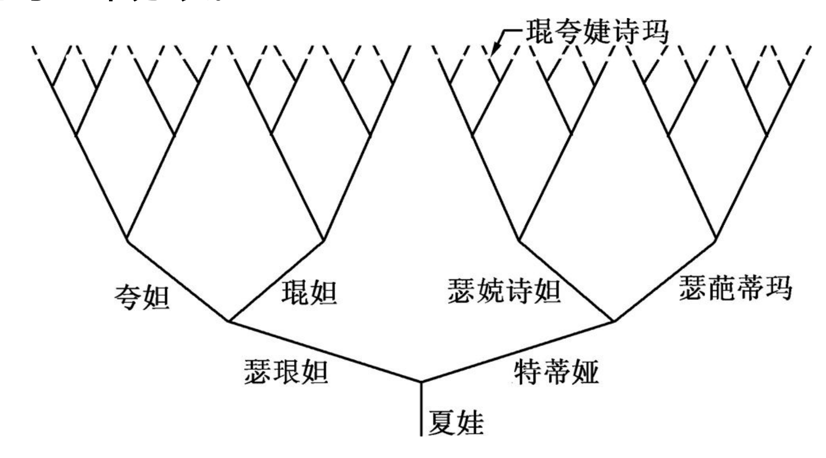
我提出下述观点。我与那个作为结果的人之间的关系对于我来说的价值，既依赖于（1）我与这个人的联系程度，又依赖于（2）这个人的物理的和心理的特征在我看来的价值。假设催眠导致我失去5种人所不欲的特征：不够整洁、过于懒惰、害怕飞行、嗜烟如命，以及悲惨的生活记忆。在此尽管远不及完全的心理联系性，却远不止是被那些坏特征的消除所超过。

我们中间很少有人认为自己是完美的。我们中的绝大多数人会欢迎在我们的物理的和精神的特征方面出现一些变化。如果出现的变化是些改进的话，我们会欢迎对两种联系都有部分的降低。如果聚合可预见地会涉及减少我所珍视的特征，增加我所厌恶的特征的话，那么我应该避免聚合。假设只有两件事赋予我的生命以意义：其一是我为社会主义而奋斗，其二是我在瓦格纳身上所发现的那些品性。如果是这样，我应该害怕与一个瓦格纳痛恨的保守分子相聚合。既然那个作为结果的人会是一个不辨音调与立场动摇的投票人，那么我与他之间的关系可能犹如死亡一样糟。但是另一种涉及到同样多的变化的聚合，我或许会当作比通常的幸存还要好。我可能把所有的变化都当作改进来看待。也许它们不是增加我求之不得的特征就是清除我所引以为憾的特征。聚合就像结婚一样，不是巨大的成功，就是巨大的灾难。

下面考虑一些更具想象色彩的人们。除了他们的繁殖方法不同之外，这些人就像我们一样。就像阿米巴虫一样，这些人是通过一个自然的分裂过程来繁殖的。这些人的生活可能就像下面的树形图表所示的那样。

这个图表上的线条代表这些人的躯体所遵循的时空轨迹。我称两个分化点之间的单线为枝。整个结构就是树。每一个树枝对应于被当作一个人的生命的东西。第一个人是夏娃。接下来的两个人是瑟琨妲和特蒂娅。沿着树到第五十个人是琨夸婕诗玛。

在他们的生命之始，瑟琨妲和特蒂娅在心理上与分裂之前的夏娃是完全相联系的。正如我已经论证的，夏娃与这两个人中的每一个的关系大约像通常的幸存一样好。相同的一些主张适用于这个共同体的历史中的其他每一个分裂。



关于夏娃与树上诸如琨夸婕诗玛那样更远的一些人们的关系我们应该如何看待？

夏娃与琨夸婕诗玛在心理上是相连续的。在这两者之间将存在一个诸直接心理联系的连续重叠链。于是夏娃具有一些类意向，这些意向被特蒂娅所实现，而特蒂娅又有一些类意向，这些又被瑟娩诗妲所实现，如此传递下去直到琨夸婕诗玛。而且琨夸婕诗玛能够虚拟记起她的直接前辈的绝大部分生活，而她的直接前辈能够虚拟记得她自己的直接前辈的绝大部分生活，依次回溯到夏娃。

尽管琨夸婕诗玛与夏娃在心理上相连续，但是在这两者之间可能没有什么与众不同的心理联系。联系性要求直接的联系。如果这些人在其他一些方面像我们，夏娃不可能与无限长的树中的每一个人密切地相联。随着时间的推移，类记忆将减弱，然后渐渐消失。虚拟抱负

一旦实现，将被其他虚拟抱负所取代。虚拟性格特点将逐渐改变。由于这样的一些事实，如果某个人处于树的更远梢头，这个人与夏娃之间的联系将更少。如果这个人相距足够遥远，两者之间可能没有什么直接的和与众不同的心理联系。假定夏娃和琨夸婕诗玛之间的情况就是如此。

我之所以使用与众不同（distinctive）这个词，是因为会存在某些种类的直接联系。琨夸婕诗玛会从夏娃继承许多事实记忆，诸如她和其他人通过分裂进行繁殖这一事实等等。而且她会继承许多一般能力，诸如说话或游泳等等。但是她不会继承把夏娃和这个共同体中的其他绝大多数人区别开来的那些心理特征。

在夏娃和琨夸婕诗玛之间存在心理连续性，但它并不与众不同。这个事例例证了我早先提及过的一个问题。这两个关系的相对重要性是什么?

我认为这两个关系都重要。其他一些人可能认为其中的一个比另一个更加重要。但是对支持这样的一个信念的论证我一无所知。我将假定其中没有哪一个关系比另一个更加重要。（这并不是假定它们的重要性确切相等。这样的一个问题不会有一个精确的答案。）

由于它在以后将是重要的，所以我将考虑一个与之不同的观点。照此观点，联系性并不重要。只有连续性重要。如果后来将有某个在心理上与现在的我相连续的人，那么在现在的我和这个人之间没有什么直接的心理联系的话，则一点也不重要。

正如我已经说过的，联系性的某些减弱也许被作为改进来欢迎。但是我们不能够可辩护地主张，如果没有什么心理联系性的话，则不重要。首先来考虑一下记忆的重要性。如果我们的生命一直是值得过的，那么我们中的绝大多数人高度珍视我们记住许多过去经验的能力。所有这些记忆的丧失不一定毁掉记忆的连续性，后者只要求记忆的一些重叠链。假设我知道从现在起的两天我惟有的经验-记忆将是有关我明天将有的经验的。照那个刚刚陈述的观点，既然将有记忆的连

续性，这就是重要的一切。我将很快丧失过去生活的所有记忆这一点对我来说不应该重要。我们中的绝大多数人会有强烈的异议。失去所有这样的记忆会是某种我们深深遗憾的事。

接着考虑我们的愿望和意向的连续性。假设我现在爱上了某些别人。我能够不再爱所有这些人，而在心理连续性上没有任何断裂。但是我会对这些变故极为遗憾。假设我还强烈地想要达到某些目标。鉴于我具有这些强烈的愿望，我会为这些愿望被其他愿望所取代而感到遗憾。我现在必须更加关心我现在所关心的成就。既然我更加关心目前的这些愿望的实现，我会因失去这些愿望而遗憾。更加通俗地说，我想要我的生活具有某些种类的全面统一性。我并不想要它非常片断化，不想要在我的愿望和所关心的事物上有不间断的波动。这样的一些波动与完全的心理连续性是相容的，但是它们会削弱心理的联系性。这是另一类我们绝大多数人会感到遗憾的变化。

最后考虑一下性格的连续性。如果我们的性格以自然的方式发生改变，这样的连续性就将存在。但是我们绝大多数人重视我们性格的某些方面。我们将希望这些不作改变。在此我们再度要联系性而不纯是连续性。

我已经表述了三个有关我们绝大多数人为什么会拒斥心理联系性并不重要这一观点的理由。既然这些是好理由，那么看来这些评述足以驳斥这个观点。我们能够认同联系性并非是重要的一切。心理连续性也是重要的。但是我们应该拒斥那个只有连续性重要的观点。

101.相继的自我

我在上文设想了一些除其自然分裂的生殖方式之外与我们没有什么两样的人们。现在我将提出这些人会如何描述他们之间的相互关系。每一个个人都是一个自我。夏娃能够把树上任何一处的任何一个个人视作她的诸子代自我之一。这个短语蕴含指向未来的心理连续

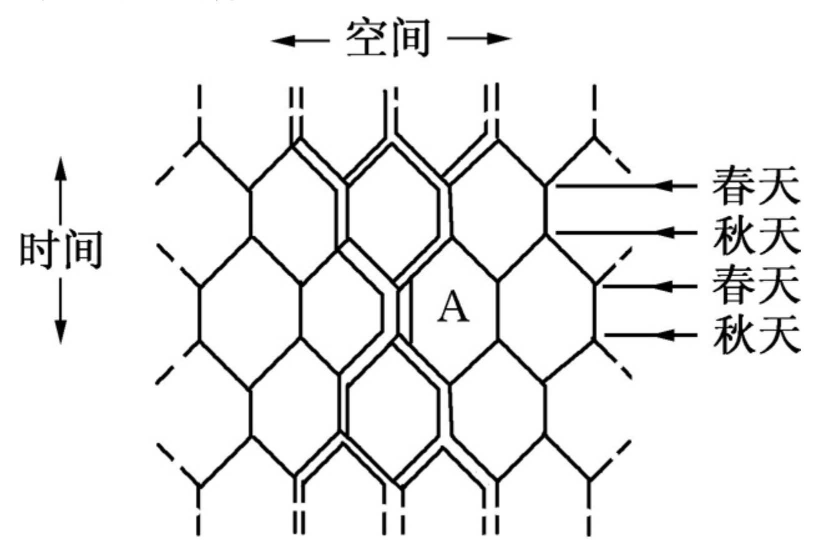
性。与夏娃不同，特蒂娅只是在树的右半边具有诸子代自我。要蕴含指向过去的连续性，这些人们能够使用祖代自我这个短语。琨夸婕诗玛的诸祖代自我就是处于把她和夏娃联系在一起的那条单线上的所有人。

既然心理连续性是一种传递关系，无论在时间指向的两个方向的哪一个方向上，即无论是……的一个祖代自我，还是……的一个子代自我，也都是传递性的。但是如果我们允许心理连续性在一个单一论证中采取这两个方向，心理连续性则不是一个传递关系。夸妲和瑟葩蒂玛两者都与夏娃在心理上相连续。但是不能由之得出他们彼此在心理上是相互连续的虚妄结论，这一结论是虚假的。

我接下来将提出这些人们会如何描述心理联系性的不同程度。我们会赋予短语我过去的自我和我未来的自我以新义。在它们的通常用法中，这些短语指称过去的或者未来的我自己。照我们的新用法，这些短语指称的不是我自己，而是那些与我具有心理联系性关系的别人。于是短语"我过去的自我之一"蕴含存在着某种程度的联系性。要蕴含与之不同的一些程度，有下列短语："我最切近的过去的自我"、"我更切近的过去的自我之一"、"我更遥远的过去的自我之一"、"几乎不是我过去的自我之一（我只能类记忆她的一些经验）"，以及"不是我过去的自我之一，只是一个祖代自我"等等。这是一系列会被琨夸婕诗玛所用的指向过去的短语。夏娃会使用一个指向未来的类似系列。

这种表达方式显然会适宜于我所设想出的人们。它能够使他们更加精确地描述处于他们之间的相互关系。这种表达方式还提供了对那个想象出的我分裂事例的一种全新的和可取的描述。尽管我并未在我的分裂之后幸存下来，但是那两个作为结果的人是我未来的自我中的两个。而且他们与我之间的密切关系就像我与明天的我自己之间那样。同样，他们每一个都能够指称我为一个同样切近的过去的自我。（他们彼此不是同一个自我，但能够共享一个过去的自我。）

下面考虑一下另一类想象中的人们。这些人们不仅通过分裂生殖而且通过聚合生殖。而且他们时常这样。他们每年秋天聚合，每年春天分裂。他们的关系如下所示：



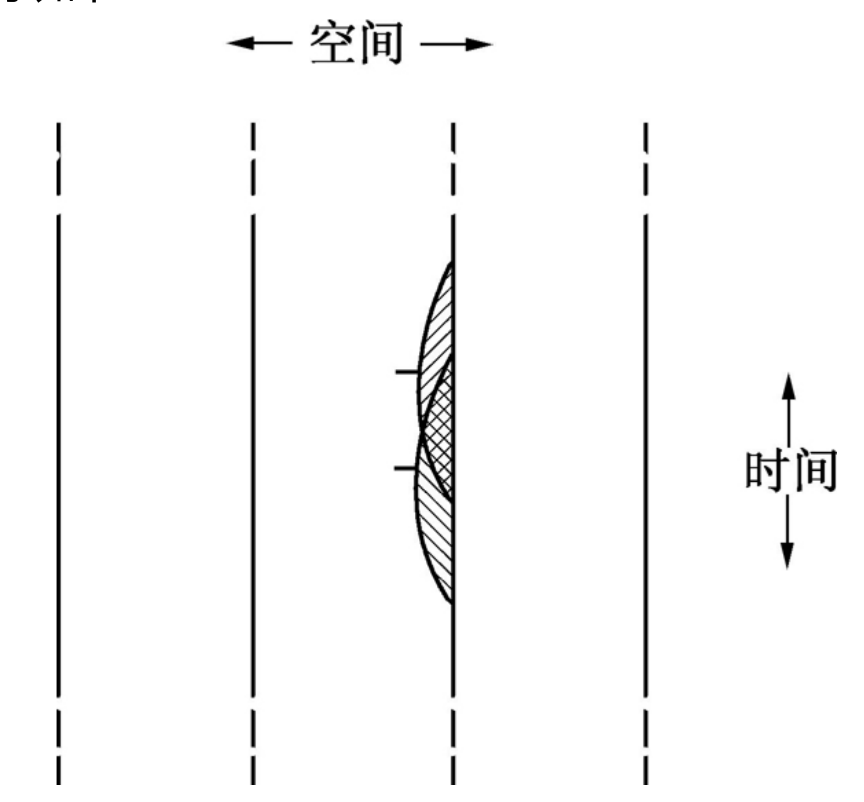
A是一个个人，其生命由那个三条线组成的支线代表。由双线组成的树形代表那些与A的生命在心理上相连续的人们。每一个人都有其双线树，它与其他人的那些树彼此重叠却不相同。

对于这些想象出来的人们而言，短语"一个祖代自我"和"一个子代自我"所涵盖的太多，以至于没有多大用处。会存在这样的一对日期，在第一个日期之前生活过的每个人都是将在第二个日期之后生活的每个人的一个祖代自我。而且既然每一个人的生命只存续半年，那么"我"这个词所涵盖的就太少，以至于这个词不能起到对我们所起到的所有作用。对于这些人们而言，要使这个词起到相当多的作用，就不得不谈论过去的和未来的诸自我。

在所建议的这种表达方式中有一个缺陷。短语"……的一个过去的自我"蕴含心理的联系性。而且这个短语的诸变体能够用来蕴含心理联系性的不同程度。但是使相继的自我区分开来的东西不是联系性的一个减弱了的程度。诸自我是由诸聚合和诸分裂来区分开来的。我们因而不能用这些短语来蕴含一个单一生命限度内减弱了的联系性。

这个缺陷并不关涉我刚刚描述的那些想象出来的人们。他们分裂和统一是如此的频繁，而且他们的生命是如此的短暂，以至于在一个单一的生命限度内心理的联系性总会保持一个非常高的程度。

最后来考虑一下另一类想象中的人们。这些人们再度只是在其生殖方式上与我们有所不同。这些人们并不进行繁殖。在他们的世界中，既没有有性生殖，也没有分裂和聚合。有的只是一些永恒的躯体，它们的外表渐渐地变化。而且直接的和与众不同的心理联系像从前一样只是处于有限的时间段中，比如像500年这样的有限时间段中。此点图示如下：



两个阴影部分代表与它们的两个中央点的心理联系性的程度。

这些人们不会用我所提出的思维方式。因为没有心理连续性的分支，他们会不得不把他们自己当作是不朽的。在某种意义上，这就是他们的实际情况。但是他们应该作出另一个区分。

这些人们会有一个理由认为他们自己是不朽的。各条“线”的那些部分在心理上是连续的。但是只有在彼此相近的部分之间才有直接的和与众不同的心理联系。这赋予这些人一个理由，不认为各个“线”对应于一个单一的、没有分化的生命。如果他们这样认为的话，他们就没有什么意指这些直接心理联系的方式了。例如，当这样的一个人说“我花了一个时期勘探喜马拉雅山”时，他的听众不会有资格假定讲话者有对于这个时期的任何记忆，或者假定他那个时候的性格与他现在的性格有任何相似之处，抑或假定他现在正在实施他那个时候具有的计划或意向。因为“我”这个词不会承载任何这些蕴含，对这些不朽的人们来说它不会具有对我们来说的那种有用性。

为赋予这些人以一种更好的表达方式，我修正我早先的一个提法。相继的自我之间的区分是能够通过诉诸心理联系性的程度而不是心理连续性的分支而作出的。由于这个联系性是一个程度问题，所以这些区分的划定就能留待说话者去选择，而且允许随着境况的不同而不同。

既然这些区分现在是在一个单一生命的限度内作出的，我们就已经更接近回归到“我过去的自我”和“我未来的自我”这些短语的通常用法了。照我所提议的表达方式，我们用“我”和其他代词只指代我们生命的这样一些部分，即当说话时我们与之具有最强烈的心理联系的那些部分。当这些联系已经显著地削弱的时候，即当性格上、生活方式上抑或信念和理想方面已经出现意义重大的变化的时候，我们可能会说：“如此这般做的不是我，而是一个早先的自我。”然后我们会描述我们以什么样的方式和在什么样的程度上与这个过去的自我相关联。

这种表达方式不仅会适合这些想象出来的不朽的人们，而且在我们自己的生活中也是经常有用的和自然而然的。这里是从两个非常不同的作者撷取的一些例子：

当我们处于爱之中的时候，我们不能充当接下去的那些个人的适当前身，当我们不再处于爱中的时候我们就已然是那些人了……[10]

我们对于未来的恐惧，对于这样一个连目睹挚爱之人的芳容、聆听所爱之人的声音的念头都必须予以断绝的未来的恐惧，对于这样一个连带给我们今天最欢乐的那些朋友都必须予以抛却的未来的恐惧，如果在这样的一种剥夺所引发的悲伤之上，我们细想起来，将再添加一种我们现在看来预计更加残酷的悲伤的话，这种恐惧非但远没驱散，反而更加强化；这更加残酷的悲伤是：根本不感到悲伤——保持冷漠，因为如果那会发生的话，那么我们的自我就已经出现了变化。在一个真实的意义上它是我们自己的死亡，这种死亡之后跟随的固然是复活，但却是在一个不同的自我之中的复活，那生活和那对生活的爱都超出了存在着的那个自我的那些注定死亡的要素所能企及的范围……[11]

不是因为其他人死了我们对他们的爱就变淡了，而是因为我们自己本人正在死去。艾伯丁没有什么理由指责她的朋友。那个正在盗用他的名字的人只是继承了它而已……我的新自我，当它从那个老自我的阴影中成长起来的时候，已然时常听到人们说起艾伯丁；通过那另一个自我……它认为它知道她，它发现她很有魅力……但这只是一种间接的爱。[12]

娜佳在她的信中写道：“当你归来……”但整个恐怖之所在是：不会有什么归来……一个新的、不熟悉的人会顶着她丈夫的名字走进来，而且她会明白那个人，那个她所爱的人，那个她洁身自爱苦等了14年的人，不复存在了……[13]

因诺肯季对她感觉有些对不起，便同意来……他感到有些对不起的不是这位曾一同生活过但这些日子没在一起生活的妻子，而是那个鬈发披肩的金发女孩，那个他在10年级时就已经认识的女孩……[14]

正如这些段落所提示的，我们情绪中的一些情绪对象可能不是没有时间考虑的另一个个人，而是这个个人生活中的某个阶段期间的另一个个人。这里是我所认为的一个常见例子。对于一对情侣来说，他们彼此相爱可能是清楚明白的。但是如果他们问彼此是否仍然处于相爱中，他们可能发现这个问题令人困惑。可能在他们看来他们仍然相爱，但是他们彼此之间的行动表现，以及他们彼此在场时的感受，看来可能并不证实这一点。如果他们在相继的自我之间作出区分，他们的困惑也许会得以消除。他们或许明白他们相爱，而且与彼此早先的自我处于相爱中。

谈论相继的自我很容易引起误解，或者说很容易被人们望文生义地理解。应该与我们细分一个国家历史的方式作一比较。我们称之为相继的国家的历史，诸如盎格鲁撒克逊的、中世纪的和都铎王朝的英格兰等等。[15]

这种表达方式还有另一个缺点。它只适宜于那些存在明显中断的事例，这样的中断标志着两个自我之间的界线。但是可能在一些减弱了的心理联系性程度中没有任何这样的中断。尽管不如同一性语言那样严格，但是谈论相继的自我不能用来表达在联系性程度上的这样的平缓减弱。在这样的事例中，我们必须直接谈论联系性的程度。

我现在将从想象性的事例转向实际生活。我将提出主张，如果我们改变关于个人同一性的性质的观点，那么这可能既改变我们有关什么是合理的信念，又改变有关什么是道德对错的信念。

---

[1]Hume（1），第269页。

[2]内格尔和威廉斯两者都倾向于认为物理连续性是重要的东西。但是内格尔认为得到一个复制的大脑会像死亡一样糟，而威廉斯认为这是现存诸种手术的一个纯粹微不足道的延伸。参见Williams（2），第47页。

[3]内格尔和威廉斯两者都倾向于认为物理连续性是重要的东西。但是内格尔认为得到一个复制的大脑会像死亡一样糟，而威廉斯认为这是现存诸种手术的一个纯粹微不足道的延伸。参见Williams（2），第44页。

[4]弗里曼·戴森在一篇著作中主张说，即便在一个无限扩张的宇宙之中，生命也会永远继续。该论证要求假定我们可以从有机躯体移居到非有机躯体。参见“无终的时间：开放宇宙之中的物理学和生物学”，《现代物理学评论》第51期（1979），第447页。

[5]Williams（2），第80—81页。另参见Brennan（1）和（2）。

[6]Quinton，见Perry（1），第66页。尽管是一个唯物主义者，但昆顿认为关系R是重要的东西，而且替代我的超然的“灵魂”。他继续说到：“对灵魂的关心完全缺位之处，根本没有什么对于个体同一性的兴趣，只有对于类的同一性的兴趣。”他写道：
在我们与其他人们的一般关系之中，他们的躯体在极大程度上并不重要。我们把它们作为方便的辨认装置使用它们，使我们能够毫无困难地定位我们所感兴趣的、我们爱好或者喜欢的持久特征和记忆综合。如果我们在感情上卷入的一个综合体结果具有一个怪异的或者令人厌恶的物理外形，那么这会是令人苦恼的；如果它不断地从一个躯体换到另一个躯体，而绝大多数这样的综合体却保持固定，那么它在社会上会是令人尴尬的；而且，假如这样的换来换去一般而言普遍的话，那么它就会是令人迷惑的和腻烦的，因为发现朋友和家人在哪里会是一件费力劳神的事情。但是，我们的关心和爱心会跟随那个特征和记忆综合而不是它原初的

躯体联合物，这一点肯定是清楚的。在普遍地换来换去的情况下，我们就会处于那些在黑暗中寻找亲人的人们所处的位置。如果那些转换频率既高空间变位又极大，我们毫无疑问将放弃辨认个体的人们的同一性，人们之间关系的整个特征就会改变，而人的生活就会像是一系列无休止的颇为短暂的海上之旅。

[7]安德烈亚·帕拉弟奥（Andrea Palladio，1508—1580），意大利建筑师。在古罗马古典主义的基础上发展出一种打破意大利文艺复兴时期装饰惯例的建筑风格。代表作品有威尼斯的罗通达别墅和基耶丽凯蒂宫。——译者

[8]乔托（Giotto，1267—1337），意大利画家、建筑师和雕刻家。被誉为意大利最伟大的前文艺复兴画家和第一位伟大的意大利文艺复兴绘画的代表。代表作多为《圣经》和圣徒题材的壁画，其中尤以帕多瓦阿雷那教堂的壁画最为著名。——译者

[9]此处所指不明，不应是德国著名作曲家理查德·瓦格纳（Richard Wagner，1813—1883），似为罗伯特·费迪南德·瓦格纳（Robert Ferdinand Wagner，1877—1953）。后者是德国出生的美国政治家，在富兰克林·D·罗斯福执政期间积极支持有关社会立法，其中最著名的当属《全国劳工关系法》（1935）。——译者

[10]Proust（2），第226页。（我对译文略有改动。）

[11]Proust（2），第349页。

[12]Proust（1），第249页。

[13]Solzhenitsyn，第232页。

[14]Solzhenitsyn，第393页。

[15]Penelhum表达了对这种说话方式的怀疑。Parfit（2）力图简短地回应这些怀疑。

第十四章

个人的同一性与合理性

102.极端主张

重新斟酌自利论。自利论宣称，对每一个人而言，有一个至高无上的合理目标：事物尽可能地适合他自己。一个理性的行动者应该既要有有利于他自己的一种时间中性的偏向，又要最终受制于这一偏向。对每一个人而言，去做他认为将会使自己变得更糟的事这一点是不合理的。

一些学者宣称，如果还原论者的观点为真，我们就没有理由关心我们自己的未来。我称此为极端主张。巴特勒写道，依照洛克观点的一个还原论者的版本，“想象……我们现在的自我会对明天将要降临到我们身上的东西感兴趣……（会是）一个谬误”[1]。从字面上来说，这是一种预测。但是巴特勒很可能真就是这么想的，即如果还原论者的观点为真，我们就没有理由表示这样的关心。在有关休谟的观点上西季威克作类似主张。照这种看法，“恒定同一的‘我’不是一个事实，而是一个虚构”；“自我只是一个……感觉系列”。西季威克问道：

> 为什么……感觉系列的一部分应……与同一系列中的另一部分，比与其他系列更有关涉呢？[2]

威金斯提出该问题无解。[3]而马代尔写道：

> 显然，如果将要处于痛苦中的人是我的话，我有一切理由去加以关心；但绝非显而易见的是，我有任何理由去关心这样一个事实，即将要处于痛苦中的人将具有某套记忆印象……

外加：

> ……并未澄清……据说在这类语境中那就是作为我所包含的一切。[4]

其他学者没有什么怀疑。照还原论者的观点，个人同一性就寓于物理的和心理的连续性。斯温伯恩宣称，如果说个人的同一性除了这些连续性就没有什么更多的东西的话，我们应该对生与死漠然置之。用他自己的话来说，“这样的连续性本身的确没有什么价值”。[5]

斯温伯恩拒斥还原论者的观点。但是至少两个还原论者作类似主张。佩里宣称，如果我仅仅知道某人将处于痛苦之中，如果能够的话，我有某个理由去防止这种痛苦。如果我获悉这个人将是我，绝大多数人会认为我有“附加理由”去防止这种痛苦。但是佩里写道，照他对个人同一性的还原论的说明，好像没有什么来证明这种主张是正当的。我有某个理由去防止一个完全陌生的人的痛苦。而且，除非它将涉及我现在的方略实现，我只具有这同一个理由来防止我自己将来的痛苦。某痛苦将是我的痛苦本身并未赋予我任何更多的理由来防止该痛苦。[6]瓦克斯贝格表示赞同。[7]

我们应该接受这种极端主张吗？我应该首先就上文所作的区分进一步加以评述。如果我们成为还原论者的话，这会怎样影响我们的态度和情绪，是一个问题；而如果还原论者的观点为真的话，这些态度或者情绪是否得到正当证明，则是另一个问题。正如我已经说过的，当我不再相信非还原论的观点时，我变得对我自己的未来不再那么关心。但是我对自己的关心仍然远远多于对一个纯粹陌生人的未来的关心。尽管我不再那么关心我的未来，但是如果我知道我以后会处于剧痛中的话，我仍然会大为悲伤。如果其他人变成还原论者，对他们的关心充其量有类似的影响而已。这是我们照任何观点所应该期待的。对自身未来的特殊关心是由进化选择的。没有这种关心的动物在把它们的基因传下去之前很可能就死掉了。即便我们判定这种关心没有被证明为正当，作为一个自然事实这种关心仍然保持着。通过努力思考那些论证，我们或许能够短暂地把这种自然关心击倒在地。但是它会很快恢复生机。

正如我业已主张的，如果某态度具有一种进化的解释，这事实是中性的。它既不支持也不削弱有关该态度获得正当证明的主张。但是有一个例外。人们可能宣称，既然我们都具有这态度，此乃认为它获得了正当证明的根据。这主张被进化的解释所削弱。既然有这解释，我们就都有这态度，即便它是没有得到正当证明的；所以我们具有这态度这一事实不能是认为它获得正当证明的一个理由。它是否获得正当证明是一个悬而未决的问题，正在等待解答。

我们应该接受极端主张，即如果还原论者的观点为真，我们就没有理由特别关心我们自己的未来吗？斟酌一下斯温伯恩作如此主张的根据。斯温伯恩宣称，物理的和心理的连续性本身没有价值。

我曾经就此相似的主张这样说：

> 这些主张过于强烈。心理连续性为什么不该有合理权重？即使照非还原论者的观点，它们也一定被赋予重要性。如果我们保持我们的同一性，但是被剥夺了所有的连续性，我们就根本做不了任何事。没有记忆和意向的联系，我们既不能行动，又不能筹划，甚至也不能思考。[8]

正如瓦克斯贝格所写的，[9]这是一个糟糕的回应。斯温伯恩认为，个人同一性是一个更深层的进一步的事实，有别于物理和心理的连续性，而且这进一步的事实赋予我们对自己的未来特别关心的理由。我一度宣称，没有心理连续性，我们既不能思考也不能行动。这并不构成对斯温伯恩观点的诘难。斯温伯恩会同意，当加上个人同一性的进一步事实时，心理连续性具有重要意义。这并不表明在该进一步事实缺失的情况下，心理连续性赋予我们特别关心的理由。

我还曾经写道：

> 当与“进一步事实”相比较时，连续性可能看起来微不足道，然而与其他任何事实比起来却弥足珍贵。所以如果根本没有进一步事实——如果它是一个幻觉——连续性可能具有至高无上的重要性。恰逢我们不是还原论者时，进一步事实看起来像太阳，照耀着我们的精神天

> 空。相较之下，连续性纯粹像白天的月亮。但当我们是还原论者时，太阳落山了。月亮现在可能比任何事物都要亮。它可能主宰着天空。[10]

斯温伯恩会拒斥这些主张。黑夜不是白昼。照斯温伯恩的观点，只有进一步事实赋予我们特别关心的理由。如果这极端主张得到正当证明，而且没有该事实，我们就根本没有这样的理由。

回到我在其中分裂的那个想象出的例子或许有所帮助。照非还原论者的观点，有3种可能性。我或许是作为分裂结果的人中的一个，或者另一个，抑或两者都不是。正如齐硕姆所写：

> 当我苦思冥想这些问题的时候，我一清二楚地看到下述事情为真……问题"我是老左吗"和"我是老右吗"有完全明确的答案。答案只是"是"和"不是"……我所要坚持的是……即便我们所有的有关个人同一性的通常标准会打破，情况也将是如此。[11]

假设齐硕姆的观点为真。非还原论者可能会说："如果我是老右，我现在有一个特别关心老右未来的理由，但是我根本没有特别关心老左未来的理由。如果我是老左，情况亦然。如果我不是两者中的任何一个，我就没有特别关心两者中任何一个的未来的理由。"

这些主张假定，在个人同一性缺失的情况下，心理连续性提供不了特别关心的理由。我们或许否定该假定。假设我将是老右。老左不是一个纯粹的陌生人。我与老左的关系现在和将来都很亲密。我们可能主张，较之一个纯粹陌生人的未来，我有理由更加关心老左的未来。

一位非还原论者可能回应说："在分裂之后，当我是老右时，老左将是确切地像我的某个别人，至少在开始时。正如你们所说的，我可能有理由后悔老左的存在。尽管我作为老右幸存下来，但是我所爱的女人并不知晓此事。对于老左错误地声称他是我，她或许会相信。不论她相信什么，老左的存在妨碍了她对我的爱。因为若此为真，所以

我会合理地希望老左快点死。既然在分裂之后我会合理地抱有此等希望，那么现在我会合理地具有此等希望。这意味，我不能有对老左的未来特别关心的理由。如果我有这样的理由的话，我一想到老左早死就该悲痛了。但是我们已经看到我会合理地欢迎这一事件。尽管老左在心理上与像我现在的我相连，但是这连续性在此并不赋予我一个特别关心的理由。”

照非还原论者的观点，如果我是作为分裂结果的人的两者之一，就能否定我有理由特别关心另一个。假如是这种余下的可能性——我不是两者中的任何一个又当如何？如果是这样的话，宣称心理的连续性赋予我特别关心的理由似乎更加令人信服。相较于纯粹的陌生人，我应该更加关心作为分裂结果的人。我有理由想要至少这两个人中的一个享其天年。相较于作为分裂结果的两人很快死掉，这对我所爱的女人会更好一些。而这个作为分裂结果的人会完成我没有完成的作品，而且以其他方式来实现我的一些愿望。

我们必须承认这种关心不同于我们所具有的对我们自己未来的关心。这是一种对于以各种方式代我而行的某个别人的关心。假如一个自我主义者有这样的关心，他可能会把这个他人当作一个纯粹的工具。假设我获悉这个作为分裂结果的人将不得不经受剧痛。我对这一消息的反应会像我获悉我将不得不经受剧痛一样吗？非还原论者们似乎会令人信服地回答不。他们可能宣称，如果这未来的痛既不透露给我所爱的女人，又不妨碍我的作品的完成，我就没有特别关心该痛的理由。这痛将不会被我感受到。如果我是一个自我主义者，我对处于痛苦中的人的关心只是这样一种关心，即这个人以各种各样的方式实现我的愿望。如果这个人的痛苦将不会妨碍到我的愿望的实现，它为什么赋予我关心的根据？

我得出结论，如果非还原论者的观点为真，非还原论者们似乎会令人信服地接受我所称的极端主张。照该主张，只有深层的进一步事

实才赋予我特别关心我的未来的理由。在这一事实缺失的情况下，心理的连续性并不赋予我这样的理由。

假设下一步一个非还原论者不再相信他的观点。他仍然会接受极端主张吗？如果个人的同一性并不包括深层的进一步事实，而只是寓于物理的和心理的连续性，我与作为结果的人中的每一个的关系就像与通常的幸存关系一样好。我与作为结果的人中的每一个人的关系是其正常原因为足够的物理连续性的关系R。如果我已经成为还原论者，我们应该接受我有关关系R像通常的幸存关系一样好的主张。但是这主张并不意味着，当R成立时，它赋予我们特别关心我们自己未来的一个理由。我们会接受这种极端主张，即如果通常的幸存并非包含深层的进一步事实，它并不赋予我们这样的一个理由。

假设我们接受我的主张中的另一个：假使心理连续性有一个反常的原因，也无关紧要。然后我们应该会同意，在电子传输中我与我的复制品的关系像通常的幸存关系一样好。但是从我们早先的非还原论者的观点的立场来看，这一主张换一种方式表达会更好。正如我曾经写到的，通常的幸存关系不比我与我的复制品的关系更好，或者说一样糟。而现在我可能说："我与我的复制品的关系并不赋予我特别关心的理由。既然通常的幸存关系不会更好，它也不赋予我这样的理由。"

这种推理思路是可辩护的。当一个非还原论者不再认为个人的同一性包含深层的进一步事实，他能够可辩护地持有他的观点，即只有这事实会赋予我们特别关心的理由。他能接受极端主张，即如果没有这样的事实，我们就没有这样的理由。而且我们会可辩护地接受这一主张，即便我们一直是还原论者。

一个非还原论者能可辩护地改变他的观点吗？他能宣称关系R赋予我们一个特别关心的理由吗？我称此为温和主张。

我相信，像极端主张一样，这个主张是可以辩护的。我不知道一个表明在这两个主张中应该接受温和主张的论证。或许可以说：

> 极端主义者们假设只有深层的进一步事实赋予我们一个特别关心的理由是错的。思考一下我们对自己的子女的特别关心，或者对任何我们所爱的人的特别关心。鉴于我们与我们子女的关系的本质，或者与我们所爱的某个人的关系的本质，我们可以令人信服地宣称我们有特别关心什么将发生在这些人身上的理由。证明这特别关心为正当的这些关系不是个人同一性的那深层的分离的事实。如果这些关系赋予我们特别关心的理由，我们可以就关系R作相同的主张。我们可以主张这关系赋予我们每一个人特别关心他自己的未来的理由。

但是一个提出极端主张者可能回应说：

> 我为什么应当关心我所爱的那些人们身上后来将发生什么？理由不可能是因为我以后将仍然爱他们。这不是答案，因为我们的问题是我为什么现在应该关心，特别是，以后我将关心什么。理由也不可能是"因为我所爱的人们现在关心以后在他们身上将发生什么"。这不是答案，因为我们的问题也是要知道为什么他们应该关心以后在他们身上将发生什么。对于在深层的进一步事实缺失的情况下，为什么我们应该特别关心我们自己的或者别的任何人的未来的问题，我们仍然没有答案。[12]

这个诘难有某种力量。而比较我们对自己未来的关心与对我们所爱的那些人的关心，可能是错误的。设若我获悉我所爱的某人很快将遭受剧痛。我将会因这消息而大为悲伤。我可能比得悉我很快将遭受此痛更悲伤。但是这关心具有不同的性质。我并不预期我所爱的某人将感受到的痛苦。或许可以宣称只有照非还原论者的观点我们才能够可以正当地预期未来的痛苦。预期或许只有由非存在的深层的进一步事实来证明为正当。大概，如果我们是还原论者，我们应该不再预期我们自己未来的痛苦。[13]

如果最后这个主张为真，它提供了认为照还原论者的观点我们没有理由特别关心我们自己的未来的一个进一步的根据。但是这个主张并不强迫我们接受这一结论。主张和否定关系R赋予我们特别关心的理由，两者看来都是可辩护的。尽管我们并非被迫接受极端主张，但

是我们可能不能够表明它应该被拒斥。在极端主张和温和主张之间存在着巨大的差异。但是我尚未找到反驳两者中的哪一个的论证。

这些结论如何涉及有关合理性的自利论？极端主张是对该理论的极端否定。这可能是针对自利论的论证，就像我在第54节中所主张的，而西季威克部分建议的。

既然极端主张是可辩护的，这论证就有某种收获。但是极端主张也是可雄辩地予以否定的。既然是这样，这论证并非反驳自利论。

103.针对自利论的一个更佳论证

我们能够做得更好。我一直在主张

（A） 既然个人的同一性并不包含深层的进一步事实，它要么不够深，要么包含得不够多。

我还辩护过

（B） 根本重要的是心理的联系性和连续性。

诉诸于（A）的极端主张能够被否定。但是（B）提供了对自利论形成新挑战的前提。该理论的核心是

同等关心的要求：一个理性的人应该对他的未来的一切的各部分同样地关心。

正如西季威克写到的，"我从今以后一年的感觉应该就像我下一分钟的感觉一样重要，只要我对它们能有同样的预测。的确，对一个人的意识生命的一切部分的这种同样的、不偏不倚的关心大概是共同的合理性观中最突出的要素……"[14]我们每一个人能够合理地对可能在较远的未来发生的事给以较少的权重，如果这种遥远性使这事件甚至较

少会发生的话。但是，照自利论，我们不能仅仅因为它更加遥远就合理地对我们较远的未来少关心一些。自利论因而主张，如果一个人知道推迟一个煎熬的到来会使煎熬更甚的话，推迟煎熬就是非理性的。

通过诉诸还原论者的观点，我们能够挑战最后的这一主张。为了简化我们的挑战，我们可假设，单单的时间的接近性无关紧要。我们可以假设，一个人仅仅因为较远之故而对较远未来予以较少关心是非理性的。这并不表明一个人对较远的未来较少关心是非理性的。这么做可能另有根据。

正如我所论证过的，根本上重要的是心理上的联系性和连续性。我还宣称过，不论联系性和连续性是怎么引起的，它们都是重要的。既然我们现在考虑的是我们实际的生命，这第二个主张就是无关的。我们的主张可以是，这两个关系就是重要的东西，倘若它们具有正常原因的话。

在第100节我还争辩说，这些关系两者都重要。我们不能可辩护地主张只有连续性重要。我们必须承认联系性重要。

既然此关系重要，我主张

（C） 我对我的未来的关心可能与现在的我和将来的我自己之间的联系程度相应。联系性是赋予我特别关心我自己的未来的理由的两个关系之一。当关心的根据之一的程度较弱时，较少的关心可以是合理的。鉴于联系性经过更长的时期几乎总是更弱，我能够合理地较少关心我较远的未来。

这一主张为一种新的贴现率辩护。这种贴现率，不是与时间本身相关，而是与根本上重要的两个关系中的一个的弱化有关。同与相对于时间的贴现率不一样的是，这种新贴现率几乎不适用在不远的将来。现在的我与明天的我之间的心理上的诸联系并不比现在的我与下个月的我之间的诸联系紧密多少。而且它们可能不比现在的我和明年的我之间的诸联系太紧密多少。但是它们比现在的我和40年后的我之间的联系大为紧密。

对（C）的辩护我一度诉诸

（D） 当某重要关系维持一个不同程度时，认为它具有不同程度的重要性并不是非理性的。

我宣称过，既然我们不能可辩护地否定（D），我们必须接受（C）。[15]正如卡根所主张的，这是另一个糟的论证。诉诸（D）并不能够支持（C）。或许可以主张

（E） 对（D）至少有一个例外。大概我们能够设想出关系的巨大弱化会为较少的关怀提供正当证明的事例。但是事实上，在正常生命中，从未有这样的一种巨大的弱化。在实际的事例中，即使联系性将维持降低了的程度，我们应该视之具有同样的重要性。仅仅因为联系性程度的这种降低而较少关心自己较远的未来，是不合理的。

如果（E）是可辩护的，那么（D）不为真。有鉴于此，我们不能借诉诸（D）而拒斥（E）。这样的诉诸以待证假定为论据。我们不能通过诉诸一个预设了某个结论的主张来为该结论论证。（举个例子可能有所帮助。假设在某个容器中放有许多玻璃球。我已经取出除了最后一粒玻璃球之外的所有玻璃球，而且这些玻璃球都是白色的。我现在取出最后一粒玻璃球，却是黑色的。我不能说："既然前面取出的都是白色的，这个玻璃球就一定是白色的。"如果对于某个一般主张存在着一个例外，我们不能只是诉诸该主张来否定该例外。）

尽管我不能诉诸（D），但是我拒斥（E）可能是对的。即便有些例外，但是当维持在一个降低了的程度时，有非常多的关系还是能够合理地被认为具有较小重要性的。其中例子有友谊、共谋、相关、受惠、近亲和负责等。相信在心理联系方面亦然，可能是可辩护的。

我主张，这信念不仅是可辩护的，而且还是不能可辩护地予以否定的。假设明天和40年后我各有痛苦的一天。在心理上我与明天的我

自己强烈地联系起来。现在的我和40年后的我自己之间的联系性要少得多。鉴于联系性是对我的未来予以关心的两个理由之一，对我来说当联系性少得多时关心少，就不能是不合理的。

既然（E）是不可辩护的，那么我们应该接受（C）。在接受（C）时，我们在拒斥处于自利论主要地位的同等关心的要求。从而应该拒斥该理论。

104.自利论者的反论证

一个自利论者现在可能诉诸

自明之理：一个人未来的一切部分是其未来的同等部分。

他可能主张，既然此为真，不对一个人未来的一切部分予以同等的关心就是不合理的。

该论证假定个人的同一性是重要的。当我们考虑设想出的我分裂的例子时，我们获悉个人同一性并非重要。尽管我将不是两个作为分裂结果的人中的任何一个，但是我与这两个人中每一个的关系包含重要的东西。

鉴于自利论者的反论证谬误地假定同一性是重要的东西，我们就不需要讨论该论证。一个前提谬误的论证说明不了什么。但是值得指出的是，即便我们认可该前提，自利论者的论证也是失败的。

（错误地）假定个人同一性是重要的东西。诉诸自明之理为同等关心的要求提供了一个好论证吗？任何人的未来的整体同样地是他的未来。这表明这人现在应该对他的整个未来同样关心吗？

如若非还原论者的观点为真的话，这或许是一个好的论证。同样，自明之理是一个深刻的真理，深得足以支持该论证。但是，照还原论者的观点，自明之理太微不足道，以至于不能支持该论证。

考虑一下

（F） 一个人的所有亲属都同等地是他的亲属。

在某种意义上，此为真。我们可以在没有程度之别的意义上使用“……的亲属”。照此用法，我的子女与我的远房堂、表兄弟姐妹同等是我的亲属。这是深层的真理吗？

必须与要求这些语词的相同用法的另一个真理区别开来。照这种用法，“……的亲属”是一个传递关系：我的亲属的亲属必定是我的亲属。这是一个有用的用法。自达尔文以来，它赋予存在巨链以全新的重要性。现在众所周知，窗外的鸟雀在某种字面意义上是我的亲属。它们在我的堂、表兄弟姐妹是我的亲属的相同意义上是我的亲属。我们具有共同的祖先。鸟雀是我远隔m辈的第n层堂、表兄弟姐妹。（“……的亲属”跨越不同物种之间的界限。如若不是的话，就不会有进化了。）

所有的高等动物确实都是我们的亲属这一点是一个深刻的真理。但是，它们都同等地是我的亲属这一点深刻地为真吗？亦即鸟雀与我的子女同样程度地是我的亲属吗？这不是一个深刻的真理。它是浮浅的，而且是让人误入歧途的，尽管事实上它从未使人误入歧途。它根本上为真这一点，是我们不得不为“……的亲属”的传递性所付出的代价。假设我们说：“我们用‘亲属’实际上指‘不太远的亲属’——远隔10辈的第10层堂、表兄弟姐妹不是真正的亲属。”这会剥夺我们关于鸟雀确实是我们的亲属这一点的深刻真理。为保持这一真理，我们必须同意——在一种浮浅的意义上——鸟雀就像我自己的子女那样地是我的亲属。

既然它是浮浅的，那么（F）不能支持我们正在考虑的这类论证。假设，由于强烈地相信亲属纽带，我把我的所有钱财留给我的各种各样的亲属。我活着时宣布我的意向，我意图把最大的份额留给我自己的子女。我的堂、表兄弟姐妹会令人信服地诉诸（F）吗？他们会争辩

说，既然他们同等地是我的亲属，他们（以及鸟雀）应该有同样的份额吗？显然不是。尽管他们同等地是我的亲属，但是该真理太微不足道，以至于不能支持他们的论证。

同理适用于

（G） 一切痛苦都同等地是痛苦。

我们能以这样的一种方式使用“痛苦”一词，这种使用方式使（G）为真。我们能论证说，因为（G）为真，更多地关心更强烈的痛苦是不合理的吗？显然不是。这是另一个太微不足道、不能支持这样的一个论证的真理。最后考虑一下

（H） 一个国家的历史的一切部分都同等地是该国家历史的各部分。

英格兰历史的一切部分都同等地是英格兰历史的部分。都铎王朝的英格兰一点不差地同样是英格兰。撒克逊人的英格兰也是如此。如果我们选择称其为“英格兰”，罗马英格兰亦然。但是，如果我们称其“罗马大不列颠”，就一点也不是英格兰了。这表明（H）微不足道。

一个国家在许多方面不像一个人。尽管有这些差异，时间中人的同一性在其根本特征上像一个时间中的国家的同一性。两者都不外乎寓于在时间中维持着各种各样的联系，其中的一些是程度问题。固然，在老年我将还是那么多地是我。但是，这真理可以公平地与（说）现代奥地利仍然还那么多地是奥地利的真理作比较。一个哈布斯堡皇室后裔会证明为正当地称这真理微不足道。

在这一节，我已经讨论了自利论者的反论证。既然该论证错误地假定个人的同一性是重要的东西，本该立刻驳回，不予考虑。但是值得表明的是，即便我们作出这个错误的假定，该论证也失败。该论证诉诸我们未来的一切部分都同等是我们未来的各部分这一主张。这个

真理太微不足道，以至于不能支持该论证。照“亲属”的一种用法，我的子女和我的堂、表兄弟姐妹都同等是我的亲属亦为真。如果我的堂、表兄弟姐妹争论说，因为这为真，他们和我的子女就应该继承同样的份额，我们应该拒斥他们的主张。同理，即使我们错误地假定个人的同一性是重要的东西，我们也应该拒斥自利论者的反论证。[16]

105.古典自利论的失败

重新回到我针对自利论的论证。我认为，该论证表明我们必须拒斥同等关心的要求。根据这一要求，我现在应该对我的未来的一切部分同等地关心。对我的较远的未来较少关心——具有经济学家们所称的一种贴现率——是非理性的。如果我具有相对于时间的贴现率，这可能是非理性的。但是如果我具有与心理联系性的程度相关的贴现率，这就不是非理性的。

自利论者可能修正他的观点。照修正论，一个理性的人的占主宰地位的关心应该是他自己的未来，但是他可能现在对他现在与之联系较不紧密的他的那些未来部分较少关心。这个修正论吸纳了我的新贴现率。照此理论，我们并非被合理地要求具有这贴现率。但是，如果我们具有，我们并非是非理性的。

这一修正造成很大变化。它打破了自利论和什么最符合一个人自己的利益之间的链接。照未经修正的或曰古典理论，对任何人而言，去做他认为将对他变得更糟的事是非理性的。照修正的自利论，必须放弃该主张。如果对一个人未来的某些部分较少关心并非非理性的话，去做他认为将对他变得更糟的事可能不是非理性的。在明知的情况下，一个人采取有悖于自己利益的行动可能并不是非理性的。

正如这最后的主张所表明的，修正论并非是自利论的一个版本。它是批判性的当前目标论的一个版本。但是我们如何归类该理论并不重要。

重要的是我们必须放弃古典理论的中心主张。考虑一下故意的和极为轻率的行动。为了我年轻时的小快乐之故，我造成下面这一点为真，即我在晚年将受大难。例如，我可能在孩提时代开始吸烟。我知道我很可能给自己带来短寿的和痛苦的死亡。我知道我是在做对我很可能变得糟得多的事。鉴于我们必须拒斥古典理论，我们不能宣称所有这样的行动都是非理性的。

照修正论，这样的行动可能是非理性的。照此论，具有与心理联系性的程度相关的贴现率并非是非理性的。当我们因现在的小快乐之故带给自己老年巨大苦难，仅当我的贴现率太急剧时，我的行动才是非理性的。

修正论的一个弱点是，它需要解释什么造成一个贴现率太急剧。但是重要之处在于，即使这贴现率不太急剧，所有这样的行动都需要受到批判。极为轻率的行动总是悲哀的，而且时常（像吸烟的例子一样）是悲剧性的。照修正论，我们不能宣称所有这样的行动是非理性的。既然我们应该批判所有诸如此类的行动，我们就必须诉诸另一个理论。

106.轻率行动之非道德性

我们应该怎样批判极为轻率的行动？或许可说，我们只能够称这样的行动是轻率的。或许可说，即使我们不再认为轻率行动是非理性的，这仍是一种批判。

许多人会拒斥这一主张。考虑一下某人是不贞洁的这个主张。现在很多人认为，不贞洁在道德上没有什么错。而且，对于这些人而言，“不贞洁的”这个指控不再是一种批判。类似的主张适用于“轻率的”这个指控。正如“不贞洁的”表达一种道德上的诘难，“轻率的”表达一种有关合理性的诘难。这从人们用来批判轻率行动的更为常见的用词可见一斑。当某人做他知道将使他变得更糟的事情时，他会被很多人叫作“蠢货”、“白痴”、“傻瓜”或者“笨蛋”。这表明该诘难是有关非理

性的。如果我们认为一个轻率行动不是非理性的话，“轻率的”这一指控对我们中的某些人来说或许就不再是一个批判。它或许像“不贞洁的”一样变成只是一种表述。

极为轻率行动应该受到批判。我们能够给以什么类型的批判呢？或许提出，我们可以诉诸批判性的当前目标论。正如我在第52节中所写到的，我们可以主张，在我们对自己的利益的关心中，不在时间上中立是非理性的。照此版本的批判性的当前目标论，极为轻率行动是非理性的。更确切地说，除非它给他人带来巨大利益，或者实现并非非理性的某个愿望，否则它是非理性的。

这提法是失败的。在我针对自利论的最新论证中，我假设了不在时间上中立是非理性的。该论证辩护一种不是与时间相关而是与心理联系性的程度相关的贴现率。既然这联系性是我关心我的未来的两个理由之一，当联系性较少时关心较少对我来说就不是非理性的。批判性的当前目标论不能可辩护地否定该主张。

对极为轻率行动的诘难必须来自另一个方面。我提出，既然我们必须拒斥古典自利论，我们就应该扩大道德性所覆盖的区域。我们的道德理论应该并吞修正的自利论所舍弃了的疆域。

正如穆勒的评判者们所主张过的，纯粹地“关注自利”的行动是罕见的。如果我极为轻率，这很可能对某些他人来说是糟的。但是，如果我的行动的主要后果将在我自己身上，那么我们中的绝大多数人不会把它判定为在道德上是错的。常识道德的诸旧版本的确包含着对一个人自己的某些义务。但是这些是特殊义务，诸如发展自己天赋的义务，或者保持自己的纯洁性的义务。人们很少主张，极为轻率行动在道德上是错的。很少作此主张的部分原因是好像没有作此主张的需要。当这样的行动被认为是非理性时，它们并不需要被认为是非道德的。但是，既然我们现在必须舍弃古典自利论，我们就应该拓展我们的道德理论。

这样做有两种方式。我们可以诉诸后果论。我们尤其可以诉诸一个不偏不倚的或者行动者中性的善行原则。假设因现在较少利益之故，我施加更大的负担在老年的我身上。这里我在做的是不偏不倚地看来具有更糟后果的事，或者说增加了受苦的总量。我们会宣称我的行动在道德上是错的，因为它增加了受苦的总量，即使将要受更多苦的是我。更宽泛地说，我的轻率行动是错的，因为我在造成更糟的结果。结果将只是对我变得更糟这一点决不是借口。

我们还会展开我们理论中相关行动者的部分。这部分涵盖我们对我们与之处于某些关系中的那些人们的义务，诸如我们的双亲、子女、学生、病人、客户或者我们的投票人。一个人与未来的他自己处于另外的特殊关系中，我们可以主张这创造了类似的特殊义务。

如果我们以这两种方式中的任何一种修正我们的道德观的话，对很多人来说这会是他们道德观念中的一个大的改变。这些人认为，不论一个人如何影响他自己的未来，它都不可能是一个道德问题。

如果我们把一个人的生命细分为前后相继的诸自我的生命，可能更容易相信这一点。正如我已经主张的，长久以来每当有心理联系性的某种显著弱化，这看起来就是自然的。在这样的一种弱化之后，我的较早的自我可能看起来与现在的我是异己的。如果我与那个较早的自我认同失败，我就在某些方面正把那个自我当作一个不同的人。

我们可以就我们自己未来的诸自我作类似的主张。如果我们现在对较远未来的我们自己几乎不关心，我们未来的诸自我就像未来的几代人一样。我们能够从更坏处影响他们，而且因为他们现在不存在，他们不能为自己辩护。像未来的几代人一样，未来的诸自我没有投票权，所以他们的利益需要特别予以保护。

重新考虑一下一个开始吸烟的男孩的事例，他知道吸烟可能导致他50年后受大苦，却漠然置之。这个男孩并不认同他未来的自我。他对这个未来的自我的态度在某些方面就像他对其他人的态度一样。这个类比使人们更加容易相信他的行动在道德上是错的。他是在冒一种

折损寿命和痛苦死亡的风险。我们应该主张，把这样的死法的风险施加在任何人身上，包括施加在这样的一个未来的自我身上，都是错的。更宽泛些说，我们应该主张，极为轻率行动在道德上是错误的。如果一些事施于他人会是错的，那么我们也不应该把这些事施于我们自己未来的诸自我。

---

[1]见Perry（1），第102页。

[2]Sidgwick（1），第418—419页。

[3]Wiggins（6）。

[4]Madell，第110页。

[5]Swinburne，第246页。

[6]“是同一的之重要性”，见Rorty，第78—85页。

[7]《个人同一性、人的性质和伦理理论》，博士论文，普林斯顿大学出版社，1983年。

[8]Parfit（6），第229页。

[9]Parfit（6），第229页。

[10]Parfit（6），第230页。

[11]Chisholm，第188—189页。

[12]这个回应是由J.Broome向我提出的。

[13]这是由Wachsberg提出的。

[14]Sidgwick（1），第124页。

[15]Parfit（6），第232页。（我的（D）版本在一个微不足道的方面属于劣等。）

[16]（该注释为1985年所加）正如B·加雷特（B.Garrett）已经指出的那样，此节有错。就亲缘而论，重要的不是“……的亲属”的传递关系，而是能够具有程度“……的近亲”之非传递关系。如果我们假定重要的是个人同一性，那么这因而就不是一个好的类比。要回答该反论证，我们可能不得不挑战这个假定。

第十五章

个人同一性与道德性

如果我们变成还原论者，那么我们在我们的道德观方面应当作一些别的改变吗？

107.自主权与家长制

当我们指使某人按照他自己的利益行动的时候，我们就是些搞家长制的人。如果我们是在阻止某个人做出非理性的行动，当其中涉及到强迫或者涉及到侵害这个人的自主权时，这是为家长制提供某种正当证明。如果我们接受自利论，上述这点就正是我们信以为自己在做的。我论证过，我们必须拒斥该理论。但我们又应该展开我们的道德理论，以便涵盖我们已经拒斥的理论。我们应该主张，极为轻率行动在道德上是错误的。

这个主张对家长式干预具有强化作用。我们强迫的那个人或许说："我可能是在非理性地行动。但是即便如此，这也是我自己的事。如果我损害的只是我自己，我就有权非理性地采取行动，而你无权阻止我。"这个反驳具有某些力量。我们的确并不认为自己具有阻止人们非理性地采取行动的一般性权利。但是我们确实认为自己具有防止人们错误地采取行动的一般性权利。这个主张可能并不适用于小的过错。但是我们认为，阻止别人犯严重错误本身非但不为过，反而会是我们经常的职责。既然我们应当认为极为轻率行动是严重错误的，那么我们就应当认为自己应该阻止这样的轻率行动的发生，即便其中涉及到强迫。自主权并不包括无缘无故地自我戕害的权利。如果一些事施于他人会是错的，那么我们就应该阻止有人把这些事施于他未来的自我。

虽然这些主张支持家长制，但是针对家长制的一些著名诘难仍然存在。如果我们每一个人都能够从自己的错误中汲取经验教训固然再

好不过，但是要让他人知道这是些错误却难上加难。

108.两种生命归宿

我们关于个人同一性的观点一旦若有改变，我们在道德观点上的转变则可能因而获得正当性证明的情形还有许多。例子之一就是我们关于堕胎的道德性的观点。照非还原论者的观点，既然我的存在要么全有要么全无，那么必定有过那么一个我开始存在的时刻。就像在我想象出的那些谱系中的事例一样，必定有一条明显的界线。有关诞生是这条界线的主张并不可取，人们也不能令人信服地在妊娠期的某处划出这条线来。如此这般，我们可能被引向这样一种观点，即认为在胎儿形成的那一刻我开始存在。我们可能主张我的生命始于那一刻。而且，按照非还原论者的观点，一个深刻的真理是：我的生命的所有部分都同等地是我的生命部分。即便我的生命刚刚开始，我同样是我。在这个时候杀死我就是直截了当地杀死了一个无辜的人。如果这是我们所相信的，那么我们将令人信服地主张堕胎在道德上是错误的。

按照还原论者的观点，我们并不认为在每一时刻我们要么存在要么不存在。我们现在能够否定一个受精卵是一个个人或曰一个人了。这就像令人信服地否定一个橡果是一棵橡树一样。在正确的条件下，一个橡果慢慢地变成一棵橡树。这种过渡不仅需要假以时日，而且是一个程度问题。没有什么明确的界线。关于个人或曰人，我们应当持同一主张。从而在有关堕胎的道德性方面，我们能够令人信服地秉持不同的观点。我们可以认为，早期堕胎没有什么不对，但是在妊娠末期堕掉一个孩子会是严重错误的。这样的一个孩子如果自己不想要的话，则该当生下来，让人领养。我们可以把介于两种情况之间的那些情况当作一些程度问题来对待。受精卵最初不是人、不是个人，但慢慢地变成人、变成个人。同样，毁掉这个有机体最初并不是严重的错误，但是慢慢地变成了严重的错误。在毁掉这个有机体绝无过错这种

情况之后，毁掉这个有机体变成一个小的过错，它只有在这个孩子后来的出生会使这个孩子的双亲或者别人的有关状况严重恶化的情形下才具有正当性。随着这个有机体完全变成一个人或者说一个个人，原来的小的过错就会转化为一个有严重错误的行动。

我已经描述了人们在堕胎的道德性方面广泛持有的两个主要观点。其中的第一个观点得到关于个人本质的非还原论者的观点的支持，第二个观点则得到还原论者的观点的支持。尽管第二个观点并非是惟一与还原论相容的观点，但是我认为我们应当采纳这第二个观点。

在这个观点范围内部，仍有分歧空间。我们中的绝大多数人并不把一个个的个人（persons）与一个个的人类一员（human beings）加以区分。但是我们中的有些人，追随洛克遗风，在这两者之间作出一种区分。这些人的典型主张是，人类的一员只有当变得具有自我意识的时候才变成一个个人。一个胎儿在妊娠期终结之前变成人类的一员。但是一个新生儿并不具有自我意识。如果我们划下这条区分线的话，我们可能认为，杀死人类的一员固然是坏的，可杀死一个个人则更坏。我们甚至可能认为，只有杀死个人才是错误的。对于这个辩论的来龙去脉我将不予追踪。至于双方的假定，我将在第四编中提出疑问。

考虑一下生命的另一个归宿。照非还原论者的观点，任何个人要么是活的要么是死的。照还原论者的观点，一个个人在其心脏停止跳动之前，会逐渐地停止存在。如果一个人精神生活的与众不同的特征逐渐消失的话，情形就属于此类。这种情况时有发生。我们能够令人信服地主张，如果那个个人已经停止存在，我们就没有什么道德理由帮助他的心脏继续跳动下去，或者说没有什么道德理由忍住不去防止这么做。

这个主张把个人与人类的一员区分开来。如果我们知道人类的一员处于不可治愈的昏迷中，就是说这个人类的一员肯定永远不能重获意识，那么我们将认定那个个人已经停止了存在。既然有一个属于人类的活体，那人类的一员则仍然存在。但是，在生命的这个归宿之处，我们应当主张只有杀死个人才是错误的。

109.该当赏罚

有些学者认为，如果还原论者的观点为真，我们就不该为我们的罪行受到惩罚。巴特勒写道，照洛克观点的一个还原论者的版本，"以我们做过的事来指控我们目前的自我（会是）一个在我们自己身上所犯下的一个谬误……"。[1]另一位18世纪的洛克批评者持有一种更笼统的主张。雷德对个人的同一性与诸如船舶或者树木一类事物的同一性进行了对照。雷德写道，这样的一些事物的同一性

> 不是完善的同一性；它毋宁说是某种我们为了说话方便之故而姑妄称之为同一性的东西。倘若变化是渐进性的，它允许主体出现巨大变化；有时甚至允许完全的改变。在通俗语言中得以与同一性相容的那些变化，不同于据认为不是在质上而是在数目上和程度上毁坏同一性的那些变化。当应用于物体的时候，同一性没有什么固定的本质；而且关于一个物体的同一性的那些问题经常是些关于语词的问题。但是，当应用于那些个人的时候，同一性没有模棱两可性，而且不允许有或多或少的程度。它是一切权利和义务的基础，是一切责任的基础；而且它的概念是固定的和精确的。[2]

雷德在关于物体的同一性方面明显是一个还原论者。照其观点，个人的同一性与物体的同一性大为不同。它涉及到一个总是具有确定性的事实，而且这事实是要么全有要么全无的。这是我所称的那个深层的进一步事实。雷德认定，这个事实是道德的基础：即正如我已经论证的，如果根本没有这样的事实，我们就不能为过去的罪行"负责"。不仅如此，所有的权利和义务的基础也都受到削弱。

某些现代学者秉持一些相似的主张。马代尔主张，“依据心理连续性所作的个人同一性分析……对我们的整套正常道德态度具有彻底的毁灭性……廉耻、自责、自豪和感激”全都系于对这种分析所得观点的拒斥。[3]而且哈克萨尔主张，我的观点削弱了一切“人权”和“非功利主义的道德约束”的基础，而且“与任何种类的人道主义道德并不相容”[4]。

我们应当接受这些极端主张吗？我们应当首先注意到下述情形。如果关于个人同一性的真理具有这些蕴含的话，我们中的绝大多数人会对此深感不安。可能有想法认为，如果这些蕴含着还原论者的观点的话，那就表明该观点是虚妄的了。并非如此。真理可能是令人不安的。考虑一下宇宙并不是由一位仁慈的上帝所创造这一主张。许多人发觉这个主张令人不安；但是这并不能够表明它是虚妄的。如果某个真理是令人不安的，那么这绝不是不去相信它的理由。它只能够是以某些方式采取行动的理由。它或许是努力对别人隐藏这个真理的理由。它甚至或许是我们为不再相信这个真理而努力欺骗我们自己的理由。正如我已经说过的，充满渴望的想法在理论上是非理性的，但在实践上可能是合理性的。例如，倘若它是我们自己从严重的压抑中解脱出来的惟一方式的话，情形就或许如此。

下面考虑一下这些极端主张中的一个。一些人论证说，如果还原论者的观点为真，我们不应当为我们的罪行受到惩罚。这个论证假定只有那个深层的进一步事实才承担应得的赏罚或者责任。我将再次问两个问题：（1）如果非还原论者的观点为真的话，这个假定令人信服吗？抑或至少是可辩护的吗？（2）倘若已知还原论者的观点为真，其假定令人信服吗？或者说可辩护吗？

回到我所想象出的分裂事例将有所帮助。照非还原论者的观点，有3个可能性，其中所有的可能性都或许是真理：我将不是两个作为

结果的人中的任何一个，或者我将是老左，抑或我将是老右，所有这一切都或许为真。假设我将是老右。老左会因我分裂前所犯下的罪行该当受罚吗？一个非还原论者会可辩护地回答道：

> 否。老左既在物理上又在心理上与分裂之前的你相连续。但是他并未犯下你的罪行。他怎么能够该当为他还不存在的时候某个别人所犯下的罪行受到惩罚呢？只有个人同一性的进一步事实承担过去罪行的责任。在这个事实阙如的情况下，物理的连续性和心理的连续性并不承担这样的责任。

非还原论者应当承认心理连续性具有某些道德蕴含。假设我过去的罪行表明我是一个杀人狂。既然他与我在心理上相连续，老左也会是一个杀人狂。这可能为在他实施犯罪之前就预防性地对他予以拘留提供了正当性证明。但是这并不表明他该当为我所犯下的罪行受到惩罚。

现在假设非还原论者改变了其有关个人同一性性质的观点。如果他变成一个还原论者，他能够可辩护地主张，我们不该为我们的罪行受到惩罚。照他的观点，该当的赏罚要求那个深层的进一步事实。既然没有这样的事实，故没有这样的该当赏罚。

就像以前的情况那样，尽管这个极端主张是可辩护的，但它也是能够可辩护地予以否定的。提出这个类比可能有所帮助。关于该当的赏罚和决定论有两种观点。照兼容者的观点，该当赏罚所要求的那种自由意志的根基不会受到决定论真理的削弱。照非兼容者的观点，决定论既削弱了自由意志的根基又削弱了该当赏罚的根基。照这第二种观点，如果我犯下我的罪行从因果性的角度来说最终不可避免的话，那么我就不该当受到惩罚。如果从道德上说把我投入监狱具有正当性的话，这也只能基于功利主义的一些根据。这样的根据之一是，我深陷囹圄可能吓阻别人实施犯罪。众所周知，如果我纯粹是被误认为犯了罪，这样的根据就会是毫不相干的。既然，甚至有罪都并不当罚，那么如果我事实上清白，就不会造成什么道德上的差异。

一些人主张，只有兼容者的观点是可辩护的。另一些人则就非兼容者的观点作此主张。还有第三种人认定，这个歧见尚未决定性地得到解决。这些人或许主张："虽然这些主张相互矛盾，从而不能两者皆真，但两者皆是可辩护的。尚未有人贡献出一个这样的论证，它既驳倒两者中的一个观点又确立起另一个观点。"

就我刚刚描述过的那两种不同观点我持此主张。非还原论者认为，个人同一性涉及到不同于物理的和心理的连续性的一个深层的进一步事实。这个主张是一个可辩护的主张：只有这个事实承担因过去罪行的该当惩罚；如果没有这样的事实，就没有什么该当的惩罚。这是非兼容者观点的一个类比。该当惩罚能够被认为与还原论不相容。但是一个与之不同的观点也是可辩护的。我们能够雄辩地主张，心理连续性随身携带因过去的罪行该当的惩罚。也许有一个决定性地化解这个歧见的论证。但是我迄今尚未发现这样的一个论证。

考虑一下下面这样一个事实：心理联系性存在着程度差异。假设，在现在的某个囚犯与犯下罪行时的他自己之间，只存在弱心理联系。通常只有在某个人犯下罪行，但事隔多年之后才被定罪的情况下，才会如此。但是，当有某种巨大的中断存在的时候，诸如当一个讲求快乐的意大利青年皈依方济各会[5]的时候，也许会同样如此。通过称囚犯为那个犯罪人后来的自我，我们能够蕴含心理联系的弱度。

有两个拘留他的根据不会受到影响。某个囚犯应当予以改造，还是应当予以预防性地拘留，要视其目前的状态而定，而不是视其与那个犯罪人之间的关系。至于第三个根据——吓阻作用，则要视另一个不同问题而定。潜在的犯罪人在意这样的一些自我吗？例如，当他们预期不会被捉住判许多年的话，他们在意吗？他们如果在意的话，吓阻了他们后来的自我，就会吓阻别人。

这会是该当的吗？洛克认为，如果我们忘记了我们所犯的那些罪行，我们就不应当受到惩罚。吉奇称这个观点是"在道德上令人厌恶

的”。[6]而且，纯粹不过是丧失记忆这么一点，看起来的确不够充分。性格的变化更加具有相关性。但是主体被复杂化了。有关该当惩罚的主张能够令人信服地得到众多五花八门的论证的支持。根据这些论证中的一些论证，记忆的丧失会是重要的。而且根据绝大多数人，性格的任何变化的本质和原因会是需要得到知晓的。

我将不考虑这些细节。但是我将作一种一般性的主张。当某个囚犯现在与他犯罪时的自己之间的联系不再那么紧密的话，他该当不再那么重的惩罚。如果联系非常微弱，他可能不该受到惩罚。这个主张看起来是令人信服的。它可能提供了为什么我们规定法定时效——确定在此之后人们的罪行不能予以追究的时间周期——的那些理由之一。（假设一个人90岁高龄，是诺贝尔和平奖的合法获得者之一，他坦白承认是他在20岁时的一次酒后斗殴中伤了一个警察。尽管这是一个严重的罪行，但是这个人现在可能并不该当受到惩罚。）

应当把这个主张与减免刑责这一概念区别开来。这个主张并不是诉诸精神疾患，而是把犯罪人后来的自我当作一个精神正常的共犯。正如一个人应得的惩罚同他参与一些人的共谋程度相对应一样，他现在因过去的罪行应得的惩罚与他现在的自己和他犯罪时的自己之间的心理联系程度相对应。

我们可能总想提出抗议：“但那同样是他的罪行。”的确如此。而且如果我们不是还原论者的话，这个真理会是一个很好的诘难。但是，照还原论者的观点，这个真理过于微不足道，不能驳倒我在有关减免刑责方面的主张。它就像这样一个主张，即“每一个共犯同样是共犯”。这样的一个主张不能够表明参与共谋没有程度之别。

在本节中我描述了3个观点。照极端主张，既然还原论者的观点为真，就没有什么人该当受罚。就像以前，这个主张是可辩护的，但它也是可辩护地予以否定的。我还主张，联系的削弱可能减免罪责。在我看来，这个主张较之否定它更加令人信服。

110.承诺

如果我们转向承诺，类似的主张也是适用的。照极端主张，既然还原论者的观点为真，我们就决不能被过去的承诺所约束。这个主张是可辩护的，但对它的否定也是可辩护的。主张联系的削弱会减免一个承诺的力量这一点是可取的。

对这些结论给出类似的辩护会令人感到索然无味。因而我转向一个新颖的问题，在该当惩罚事例中与之不存在任何类比性。当我们考虑承诺的时候，个人同一性的事实两度加入进来。我们必须既考虑承诺者的同一性，又必须考虑被承诺者的同一性。联系性的削弱可能减免承诺者的义务。但是，从那个得到承诺的人的立场出发，还原论者的观点的任何蕴含都会刻意地遭到杜绝。我们会要求这样形式的承诺："我将帮助你，以及你的所有后来的自我。"如果向我所作的承诺呈现为这种形式的话，它们就不能被认为后来会受到我的性格中的任何变化或者我余生的心理联系性中的任何弱化的破坏。

这里存在着一种非对称性。类似的形式并不能如此显然地约束承诺者。我可能说："我，以及所有的我后来的自我，将帮助你。"但是它会遭到诘难——我能够约束的只是我目前的自我。这个诘难具有某种力量，因为它类似于我能够约束的只是我自己这一令人信服的主张。与之形成鲜明的对比的是，没有人否定：我能够向你承诺我将帮助别人，诸如你的孩子们。因而，显然我能够向你承诺我将帮助你后来的诸自我。

这样的一个承诺可能变得特别具有约束力。假设你的改变要远远大于我的改变。那么我可能把我自己当作是作出了承诺的，但不是向你，而是向你早先的自我。因而我可能想，你不能够免除我的承诺。这就像对某个现在已经不在世的人所作的一个帮助他的子女的承诺。我们不能得以解除这样的一些承诺。

这样的情形会是罕见的。但是，因为它说明了某些别的要点，还是值得给出一个例子。考虑一下

> 19世纪的俄罗斯人。在数年之后，一个年轻的俄罗斯人将继承一大笔地产。因为他具有社会主义理想，他现在打算把这些土地分赠给农民们。但是他知道随着时间的推移他的理想会逐渐淡化。为了防范这种可能性，他做了两件事情。他首先签署了一份法律文件，这份文件将保证自动交出土地，而且这份文件只有在他妻子赞同的情况下才能够被撤销。然后他对妻子说："你要向我承诺，如果我万一改变主意，要你撤销这份文件的话，你将不予赞同。"他还补充说："我把我的理想当作我的实质性的东西。如果我失去这些理想，我想要你认为我已不存在了。我想要你把那时的你丈夫不当作我，不当作向你要求这个承诺的那个人，而只是当作他的堕落的后来的自我。向我承诺，你不会按照他的要求去做。"

这个恳求，使用了相继的自我的语言，看起来既是可理解的又是自然的。而且如果那个人的妻子作出了这个承诺，并且他的确在人到中年的时候要求他的妻子撤销那份文件的话，她或许有理由认定她自己的承诺没有被免除。可能在她看来，她就像对两个不同的人具有义务。她可能认为，做她现在的丈夫所要求的事情会是对那个她所爱和所嫁的年轻人的一种背叛。而且她可能认定她现在的丈夫所说的话不能使她免除对这个年轻人——她丈夫的早先的自我——的背弃责任。

这样的一个例子可能看起来无需在相继的自我之间作出区分。假设我要求你承诺永远不给我香烟，即使我向你乞求也不给。你可能认为我不能够，通过乞求你，就解除了你的这个承诺。而要这么想，你无需否定你所作承诺的对象是我。

确实如此。但理由是烟瘾遮蔽了判断。类似的一些例子或许涉及到极大的压力或痛楚，抑或极度的诱惑——就像奥德修斯[7]所遇到的情况那样，当海上女妖塞壬[8]唱歌的时候他把自己捆绑在桅杆上。当没有什么遮蔽一个人的判断的时候，我们中的绝大多数人相信我们向他作出承诺的那个人总是能够使我们解脱出来。他总是能够，如果处于健全的精神状态，免除我们的承诺。无论承诺可能是什么，我们都相信这一点。照这个观点，一个承诺的内容不能防止它被免除。这像

有关权威的一个类似事实。假设一位将军命令他的部队说：“我命令你们拂晓发起进攻，不要理睬任何后来的相反命令。”他后来说：“别管我过去的命令，现在撤退。”尽管有第一道命令的内容，但是他的部队应当遵守的命令会是这第二道命令。

再回到那对俄罗斯夫妇。那个年轻人的理想淡化了，而且在人到中年的时候要求妻子撤销那个文件。尽管她向他承诺过会拒绝，他宣布他现在免除她的承诺。我已经描述了两种她可能认为她没有获得解除的方式。她或许把她丈夫的改变主意，当作表明他现在不能作出考虑周详的判断来理解。但是我们能够假设她没有这样的思想。我们还能够假设她分享我们有关承诺的观点。如果是这样，她怎么能够认为她的丈夫不能免除她的承诺呢？仅当她认为在某种意义上她的承诺对象不是他的时候，她才能够相信这一点。我已经描述了这样的一种意义。她可能把那个年轻人的理想的丧失当作涉及到他被一个后来的自我所取代。

这个例子例证了一个一般性的主张。我们可能把一个人一生中的某些事件当作在某些方面就像出生或者死亡。不是在所有的方面，因为在超出这些事件之外那个人具有早先的或者后来的诸自我。但是，我们的情绪中的一些情绪的对象，以及我们的原则中的一些原则所应用的对象，可能只是诸自我的诸系列中的一个系列而已。

那个年轻的俄罗斯社会主义者把他的理想当作对他目前的自我而言是实质性的。他要求他的妻子向这个目前的自我承诺不要采取违背这些理想的行动。而且，照这种思考方式，她永远不能得以解除她的承诺。得到她的承诺的那个自我会由于试图解除她的承诺而不再存在。

此处的要点不是一个墨守成规的问题。部分上它是有关这个妇女的信念和情绪的一个真理。她爱的不是她的人到中年的丈夫，而是那个她当年所嫁的年轻人。这就是为什么她认为应该对这个年轻人忠诚的原因。我们能够爱某个亡故的人，而且能够相信我们对某个已经亡

故的人抱有承诺。而且这样的爱和承诺的对象可以不是某个亡故的人，而是某个健在的人的早先的自我。

可能有诘难说，通过以方便的方式区分相继的自我，我们会不公正地逃脱我们的承诺，或者我们应得的惩罚。并非如此。我可能说："今晨抢劫银行的不是我，只是我的过去的自我。"但是别人会更令人信服地回驳说："它是你。"既然没有什么固定的准则，我们能够选择什么时候说有一个新自我。但是这样的一些选择可能是，而且据人们所知是，非真诚的选择。而且它们还能够真诚地表达信念——那些并不是本身被选择的信念。我所举例子中的那个妇女的情形就是如此。她所爱和所嫁的那个年轻人在某种意义上已经不再存在——她的人到中年的、愤世嫉俗的丈夫充其量不过是这个年轻人的后来的自我——这些主张对她来说好像比"他们是同一人"这一简单的主张表达出了更多的真理。正如如果我们把俄罗斯历史分成帝国时代的历史和苏联时代的历史，我们能够更加精确地描述俄罗斯的历史那样，如果她把她丈夫的生命分成两个相继的自我的生命的话，她能够更加精确地描述她的丈夫的生命。[9]

111.个人的独立性和分配正义

我们是一些彼此不同的人，每一个人都有自己的生活要过。照个人同一性的性质的所有观点来说，这一点都是如此。但是照非还原论者的观点，这是一个更加深刻的真理。如果我们接受这个观点，我们可能把这个真理当作支撑一切行动理由的那些根本事实之一。这个事实被称作个人的独立性（separateness of persons）。

西季威克认为，这个事实是合理性问题上的自利论的基础。如果基本的东西是我们是些不同的个人，每一个人都有自己的生活要过，那么这就支持这样一个主张，即对每一个个人而言最合理的终极目标是他自己的生活尽可能好地进展。西季威克认为，还有另一个同等合

理的终极目标。这就是，对每一个人而言事物总体上尽可能好地运转。在道德赋予我们的终极目标这一点上，许多人同意西季威克的观点。而且一些人接受了西季威克这样的一个观点，即当道德与自利相冲突的时候，对于我们最有理由做什么这个问题没有答案。当他比较道德的理由和自利的理由的时候，在西季威克看来两者中没有哪一个超过另一个。

西季威克持有这个观点是因为他认为个人的独立性是一个深刻的真理。他认为，诉诸这个真理是赋予自利论者针对道德性主张的一个充分辩护。而且他提出，如果我们就个人同一性持一个与之不同的观点的话，我们就会驳斥自利论。我已经主张，这点为真。

自利论，在西季威克看来，似乎基于个人的独立性之上。我现在将讨论关于道德性的一个类似主张。这个主张对西季威克的道德观点形成挑战。西季威克认为，存在一个终极的道德原则，即不偏不倚的仁爱。由于他采纳关于自利的快乐论，所以他的仁爱原则呈现为一种快乐主义的形式。照他的观点，我们的终极道德目标是减去痛苦之后最大的幸福净总额，或者减去"人所不欲的意识"之后最大的"人所共欲的意识"净总额。那些拒斥快乐主义的功利主义者们把终极目标当作减去负担之后最大的利益净总额。照两者中的任何一个版本，功利主义者的观点，在下面的意义上是非人称的。所重要的一切是幸福和痛苦的数量或者利益和负担的数量。这些数量在不同的人们之间如何分配则没有什么道德区别。

许多人拒斥这个观点。他们可能说："功利主义者的目标可能是我们的终极道德目标之一。但是我们至少还有另一个目标。幸福和痛苦，或者利益和负担，应该在不同的人们之间公平地分享。在功利主义的原则之外，我们还需要一些分配正义的原则。例子之一就是平等原则。照这个原则，如果一些人的状况比别人恶化，而这却非他们自己之过，那么这是糟的。"

对平等的论证经常被人们主张基于个人的独立性之上。一个这样的主张或许是："既然我们过着不同的生活是一个深刻的真理，那么，只要我们同样是应得的，每一个人的生活都应当过得同等地好；这是一个终极的道德目标。如果这点是不可能的，那么至少每一个人的生活都应当有好好运转的同等机会，这是真的。"[10]

如果我们不再相信非还原论者的观点，关于平等原则和别的一些分配原则这蕴含什么？我的主要主张将是这样一些：这个观点转变支持关于这些原则的三个论证。这些论证中的两个蕴含着我们应当赋予这些原则以更大的应用范围（more scope）。这会使它们更加重要。但我们也可能被引至赋予这些原则以更小的权重（less weight）。这会使他们更加不重要。因而我们必须问净效果会是什么。

112.功利主义者的观点的三种解释

在提出这些论证之前，我将以下述方式提及两个相关的主张。功利主义者们拒斥分配原则。他们瞄准的是减去负担之后的最大的利益净总额，而不管其分配如何。可以说，他们追求最大化。

当我们的行动只能够影响到一个个人的时候，我们中的绝大多数人接受最大化。我们并不认为，我们应该给予某个人的幸福天数更少一些，以便更公正地把它们分摊到他的生命历程的不同部分。当然，存在一些支持对享受进行分摊的论证。我们仍旧保持清醒，而且前面有更多的可以期待。但是这些论证不能算作反对最大化；而是提醒我们如何获得它。

当我们的行动能够影响不同的一些人的时候，功利主义者们作类似的一些主张。他们接纳一些支持分摊享受的新论证，诸如诉诸相对剥夺的后果或者趋小的边际效用的那些论证。但是功利主义者们把平等当作一种纯粹的手段，不是作为一种单独的目标。

由于功利主义者们对待生命集合的态度就像我们对待单一生命的态度，所以他们忽略了不同生命之间的界线。我们可能问："为什么？"

这里是3个有关的提法：

（1）他们的道德推理方法导致他们忽视这些分界。

（2）他们认为那些分界是不重要的，因为他们认为生命的集合就像单一的生命。

（3）他们接受还原论者的关于个人同一性的观点。

提法（1）是由罗尔斯提出的。[11]它能够这样加以总括。许多功利主义者们用人们所称的不偏不倚的观察者的方法来回答道德问题。当这样的一个功利主义者问自己，作为一个观察者，什么会是正确的或者什么是他所不偏不倚地选择的东西的时候，他可能自我认同于所有受到影响的人们。他可能想象自己会是所有这些彼此不同的人们。这会引导他忽视彼此不同的人们受到了影响这一事实，以至于忽视了那些在不同的人们之间进行公正分配的主张。

提法（2）是由戈捷和另外一些人提出来的。[12]照此见解，功利主义者们必须假定人类是一种超有机物，或者像某些印度教徒所相信的，人类处于一个单一的世界灵魂之中。倘若提法（1）和（2）为真的话，它们就以削弱功利主义者的观点的方式解释了这个观点。显然，忽视我们过着不同的生活这个事实是一个错误。而且人类不是一种超有机物。

我提议提法（3）。照这个提法，功利主义者们之所以拒斥那些分配原则，是因为他们相信还原论者的观点。如果还原论者的观点支持对这些原则的拒斥，那么这第三个解释非但不削弱反而支持功利主义者的观点。

就某些功利主义者来说，提法（1）可能是正确的。很多功利主义者如他们是不偏不倚的观察者那样考虑道德问题。像罗尔斯所主张

的，这些人中的一些人可能是认同性的观察者们。但是还能够有超然的观察者们。一个认同性的观察者想象自己是所有受到影响的人们，一个罗尔斯主义者想象自己是受到影响的人之一（不知道是谁），而一个超然的观察者想象自己在那些受到影响的人中谁都不是。

一些功利主义者一直是超然的不偏不倚的观察者。这些功利主义者并不忽视人们之间的分界。而且就像罗尔斯所评述的，好像没有什么理由应当导致超然的观察者忽略那些分配正义原则。如果我们以这种超然的方式处理道德——如果我们并不把自己当作潜在地涉及到其中——我们可能更有点倾向于拒斥这些原则，因为我们不会担心自己也许会成为那些情况恶化的人们之一。但是这个处理道德问题的特定方式未能充分解释为什么这些功利主义者们拒斥那些分配原则。

提法（2）正确吗？作为对功利主义者的观点的一个解释，（2）是虚假的。黑格尔的一些追随者们认为，国家是一个超有机体。引用其中的一个学者的话来说，国家"是一个有机体，就像个人一样。"[13] 但是功利主义者忽略了国家疆界，而且他们并不认为人类是这样的一个单一存在。

与其把提法（2）当作对功利主义者的观点的一种解释，不如把提法（2）当作对这个观点的一个诘难。这个见解可能是：功利主义者的观点不能得到正当证明，除非人类是一个超有机体；而且，既然这是虚假的，功利主义者拒斥那些分配原则就是错的。

我提出一个不同的解释。照提法（3），功利主义者之所以忽略了分配，是因为他们接受了还原论者的观点。（3）与（1）是相容的。一些功利主义者可能既是些认同性的观察者，又接受还原论者的观点。但是（3）与（2）相冲突。

此处看起来可能存在一种困惑。照提法（2），人群被比作单一个个人。这是还原论者的观点的倒置；还原论者的观点把一个个人的

历史比作一个国家的历史或者一个人群的历史。既然这两个观点都把国家比作人，它们何以是不同的观点?

答案是这样的。当我们考虑国家的时候，我们中的绝大多数人是还原论者。我们认为，一个国家的存在涉及到的只不过是在其疆土上生活的、以某些方式行动的国民的存在。与之形成对照的是，当考虑个人的时候，我们中的绝大多数人相信非还原论者的观点。我们认为我们的同一性必定是确定的。除非一个个人是不同于其大脑、躯体及其经验的独立存在的实体，这才不能为真。于是，我们中的绝大多数人关于国家是还原论者，关于个人则不是。正是两个常见观点之间的这个差异解释了这两个对比。X像Y这一主张典型地假定了对Y的那个观点是常见的。因而，如果我们关于两者都是还原论者的话，我们将说："人像国家。"如果我们关于两者都是非还原论者的话，我们将转而说："国家像人。"对超有机体的信念可能是关于国家的一个非还原论者的观点。

113.变更原则的应用范围

既然功利主义者们拒斥分配原则，他们认为生命的界线没有什么道德重要性。照他们的观点，个人的独立性是可以忽略的。我们已经描述了对这个观点的3种解释。我现在将论证，尽管有某些复杂之处，但是我的解释是最好的解释。

考虑一下

儿童的负担。我们必须决定是否让某个儿童经受某些辛苦。如果我们决定这么做的话，这要么

（1）是出于这个儿童成年时的更大利益，要么

（2）是出于某个别人的类似利益——诸如这个儿童的弟弟。

（1）或者（2）是否为真在道德上重要吗?

我们中的绝大多数人会回答：“是。如果是出于这个儿童自己后来的利益，就至少不可能有不公平。”我们可能补充这样一个一般性的主张，即如果某些负担是有益于某个人自己的善的话，施加这些有益的负担更加可能得到正当性证明。

功利主义者会接受这个主张，但却以一种不同的方式加以解释。他们没有主张这样的一些负担不能是不公平的，而是主张它们一般而言更加容易忍受。

为了防堵这个回驳，我们能够假设我们的孩子太小，以至于尚不能以这种方式获得激励。这例证了歧见之所在。功利主义者会说：“把这个负担加在这个孩子身上是否对，只取决于后来的利益多么大。它并不依赖于谁受益。如果受益的不是这个孩子本人而是某个别人，在道德上不会有什么差异。”非还原论者会反驳说：“相反，如果受益者是那个孩子本人，这会有助于证明负担的正当性。如果受益者是某个别人，就会是不公平的。”

那两个关于个人同一性本质的观点在这个分歧中支持不同的方面吗？

部分答案是明确的。非功利主义者认为，受益的将是作为一个成年人的那个孩子本人是一个在道德上重要的事实。照非还原论者的观点，这个事实更加重要，因为正是照此观点这个孩子和那成年人之间的同一性就其性质而言更加深刻。照还原论者的观点，这个同一性中所牵扯的东西更少深刻性，而且经过青春期它保持一种减弱了的程度。如果我们是还原论者，我们可能把那个孩子和他的成年的自我之间的联系的弱化比作不同的人们之间联系的缺乏。我们将给予这样的一个事实以更大的权重，即在这个例子中这个孩子并不在乎他的成年的自我身上将发生什么。如此这般，受益的将是他这一点在我们看来可能不大重要。我们也许说：“受益的将不是他。受益的将只是他的成年的自我。”

非还原论者的观点支持非功利主义者的回应。能够由此得出还原论者的观点支持功利主义者的主张吗？其实不然。我们可能说："正如受益者是某个别人会不公平一样，所以如果受益者不是那个孩子，而只是他的成年的自我，也会是不公平的。"

这个论点是一个一般性的论点。如果我们是还原论者，我们把生命之内的粗略划分当作在某些方面不同生命之间的划分。因而我们可能最终会把两类分配一视同仁，即把生命之内的分配和生命之间的分配一视同仁。但是一视同仁的方式却有两种。我们可以把分配原则应用于两者，也可以把分配原则不应用于两者中的任何一个。

我们会采取其中的哪一种方式？我区分出我们的道德观点会出现变化的两种方式。我们会赋予分配原则以一个不同的应用范围和不同的权重。如果我们成为还原论者，我们会被引导到赋予这些原则以更大的应用范围。既然我们把生命之内的划分当作在某些方面像生命之间的划分，那么甚至把分配原则应用于生命之内都是可能的，就像把负担施加在一个孩子身上的那个主张中的情况一样。通过拓宽分配原则的应用范围，我们会进一步脱离功利主义者的观点。在这个方面，还原论者的观点不是有利于功利主义者的观点，而是不利于功利主义者的观点。

114.变更原则的权重

下面转向功利主义的观点的第二种解释。戈捷提出，假设我们应当为人类追求最大化，"就是假设人类是一个超级个人（super-person）"。[14]

为了理解这个提法，我们应当首先问为什么在一个单一生命中我们能够忽略分配原则。只追求最大化在此处为什么在道德上是允许的？可能会有人认为这之所以是允许的，是因为我们如何对待我们自己的生命不是一个道德问题。即便这为真，确当的解释也不会是这样。我们认为在某个别人的生命中最大化能够是对的。医学提供了一

些例子。我们认为，对医生们而言，代表他们的处于无意识中的病人实现最大化是对的。他们选择某种给病人造成较小痛苦总量的手术是对的，即便这种痛苦都会出现在一个时段之内。我们并不认为这对于这个时段的这个人会是不公平的。

一些人主张："我们有在一个生命之内最大化的自由，只因为它是一个生命。"这个主张支持戈捷对功利主义者们的指控。它支持这样一种主张，即只有当不同的生命像一个单一生命的不同部分的时候，关于不同的生命我们才会自由地最大化。

当面对这个论证的时候，功利主义者们会否定它的前提。他们可能主张："为最大化提供正当性证明的不是一个生命的统一性。受苦是坏的，而幸福是好的。如果坏的东西少一些，而好的东西多一些，则更好。这足以证明最大化的正当性。既然不是一个生命的统一性在这个生命之内证明最大化的正当性，这能够在不假设人类是一个超级个人的情况下，经过不同的生命得到正当性证明。"

与还原论者的观点的一个联系就是这一点。正是照这一点，而不是照非还原论者的观点，戈捷的论证是更加令人信服地予以否定的。如果一个生命的统一性是较少深刻的，主张这个统一性并不是正当证明最大化的东西就是更加令人信服的。这是还原论者的观点为功利主义者的观点提供某种支持的方式之一。

我将展开这些评述。有两类分配：生命之内的分配和生命之间的分配。而且有两种把它们一视同仁的方式。我们能够把分配原则应用于两者，或者不把分配原则应用于两者中的任何一个。

功利主义者们不把这些分配原则应用于两者中的任何一个。我提示，这可能部分地是因为他们接受了还原论者的观点。一个不相容的提法是，他们接受了相反的观点，相信人类是一个超级个人。

如果我们忽视了有两条路线都通往舍弃分配原则这一事实的话，我的提示可能看起来明显是错的。我们可能不给予它们什么发挥的余地，或者不给予它们什么权重。

假设我们假定通往舍弃分配原则的惟一的路线是其应用范围的改变。这是由罗尔斯的主张——“功利主义者把适合于一个人的选择原则展开到社会”——所提出的。[15]此处的假定是，通往功利主义的路线是应用范围中的一个变化，不是分配原则的变化，而是其关连物的变化：我们忽略这些原则的自由。如果我们假定惟一的路线是应用范围的变化，的确可能看起来功利主义者们必定要么假定任何人群都像一个单一的个人（戈捷的提法），要么至少忘记了并非如此（罗尔斯的提法）。

我将描述另一条路线。功利主义者们可能并不是在否定那些分配原则具有应用范围。他们可能是在否定它们具有权重。这否定可能得到还原论者的观点的某些支持。

更确切点说，我的提法是这样的。还原论者的观点的确支持分配原则的应用范围的变化。它支持给予这些原则更大的应用范围，以便它们甚至应用于一个单一的生命之内。这就是我在“儿童的负担”这个事例中所主张的。一个还原论者更有可能把这个孩子与其成年的自我的关系当作像与一个不同的个人之间的关系。于是，他更可能主张只是使他的成年的自我受益而对这个孩子施加负担是不公平的。正是照非还原论者的观点，我们能够更令人信服地回应说：“这不能是不公平的，因为后来将受益的人同样是他。”像我们后面将看到的，还有另一个论证；照还原论者的观点，这个论证支持分配原则的应用范围的更大展开。但是，尽管还原论者的观点以这两种方式支持展开分配原则的范围，但是它也支持给予这些原则更少的权重。而且，如果我们没有给予这些原则权重，那么，我们已经赋予它们更广的范围这一点将并不会造成什么不同。这就是净效果何以可能是功利主义者的观点之所在。

这个提示在下述方面与一些别的提示不同。罗尔斯评述道，功利主义者的观点看来涉及到“把所有的人融合为一个人”[16]。内格尔同样主张，功利主义者“把彼此清晰可辨的诸个人的……愿望当作它们是一

个大规模的个人的……愿望”[17]。而且我已经援引过戈捷的类似主张。照我的提议，功利主义者的观点得到的可能不是诸个人的融合的支持，而是他们的部分瓦解的支持。它可能依靠这样一个观点，即一个个人的生命比我们绝大多数人所假定的要更少深刻地整合在一道。功利主义者们对待利益和负担，可能不是把它们当作来自同一生命之内，而是当作它们无论来自何处都没有什么道德差异。而且这个信念可能部分地得到这样的观点的支持，即每一个生命的统一性以及从而不同生命之间的差异性，就其性质而言是不太深刻的。

由于忽视不同的人们之间的分配原则，功利主义者的观点是非人称的。罗尔斯提出，它“错把非人称性当作不偏不倚”。[18]如果功利主义者们努力做到以不偏不倚的方式引导他们忽视不同的个人之间的差异的话，就会如此。这可能是对于那些功利主义者们所作出的主张，这些功利主义者们的道德推理方法的确具有这个后果。它可能是关于一个“认同性的不偏不倚的观察者”所作出的主张，其道德推理方法导致他想象他本人将是所有受到影响的人们。但是迄今鲜有功利主义者们以这种方式进行推理。而且，照我的提法，功利主义者们无需错把非人称性当作不偏不倚。他们的非人称性可能部分地得到关于个人性质的还原论者的观点的支持。正如罗尔斯所写的：“适用于任何事物的正确管理原则依赖于那个事物的性质。”[19]

115.使一个人承受负担只是要使某个别人受益这一点能对吗？

现在我将展开我的提议。功利主义者们认为，即便利益和负担涉及到不同的人，也可以在利益和负担之间径直进行彼此权衡。此点通常遭到否定。

我们首先可以区分两种权衡。声言某个负担事实上重于另一个负担，是声言这一负担更大。声言某个负担道德上重于另一个负担，是声言我们应该解除这一负担，即便代价是没能解除另一个负担也在所不惜。不同负担之间的权衡以及负担与利益之间的权衡亦然。某个利

益何以能够大于某个负担，或者某个利益何以事实上重于某个负担，值得予以解释。如果当人们面对要么利益和负担两者皆选要么两者都不选的选择的时候，人人会选择两者皆选的话，这就再清楚不过了。此处每一个人都会认为，因这个利益之故值得忍受这个负担。这是一个相当好的检验标准，人们必须对他们的生活中的诸部分给予同等的关注，他们正是在这些生活的一个个部分中接受这些利益和负担的。鉴于绝大多数人对较遥远的未来关心较少，通过问人们是否会选择在得到这个利益之前承受这个负担这样一个问题，该检验标准得到最好的应用。如果他们认为这值得做，这就表明在他们那里这个利益事实上重于这个负担。

某些人主张，如果两个负担出现在不同的生活中，一个负担就不可能事实上重于另一个负担。这些人主张，人与人之间的这种比较毫无意义。如果我失去了我的手指，你失去了你的性命，主张你的损失大于我的损失毫无意义。对于这一观点，我在此处存而不论。

另一些人主张，不同的人的生活中的负担和利益不能从道德上加以权衡。我将对这个观点的一个部分予以考虑。这是这样一个主张，即某个人的负担在道德上不能纯然被某个别的他人的利益在权重上超过。我说纯然利益，是因为该主张并不意在否定给某个人负担以便令他人受益这样一点可能是对的。这也许是分配正义所要求的。我们可以正当地向富人课税，以使穷人受益。该主张所否定的是这样的行动只有基于利益大于负担的功利主义的根据才能够得到正当性的证明这一点上。

这个主张经常以一些限定形式出现。这个主张可以局限于一些巨大的负担，或者，呈现为这样一种主张，即，若使某个人的负担超重，那么他人的受益必须大得多得多。在这里我将讨论这个主张的最简单的一种形式，因为我的绝大多数评述也适用于其他形式。罗尔斯对该主张的表述为："平衡不同的人之得失的推理……被排除了。"[20]我称此为对平衡的诘难。

这个诘难部分地基于另一个不同的主张。这个主张是，一个人的负担不能被他人的受益补偿。我称之为有关补偿的主张。加诸一个限定条件，这个主张显然是成立的。在某种意义上，我们的那些负担能够被我们所爱的那些人的受益而补偿。但是它们不能被其他人的受益而补偿。

有关补偿的主张不能被否定。如果成为还原论者影响到我们有关这个主张的观点，那么后果就是这些。我们或许首先甚至在单一的人的诸生活中展开这个主张。因此，在我所给出的例子中，我们或许主张那个孩子的负担不能由他的成年的自我的受益而得到补偿。或者我们或许主张，这里不可能有完全的补偿。这或许支持这样一种主张，即仅当他的成年的自我的受益大得多的时候，这个孩子的负担在道德上才能够被超过。这些主张就像那样一些主张——当心理联系已经大为削弱的时候，我们因自己早先的那些自我的所作所为而该当承受的惩罚或者该当承担的责任要小一些。在某些方面或者在某种程度上，那些主张把一个生活中联系弱的那些部分当作不同的生活来对待。这样的一些主张因而变更了我们的那些原则的应用范围。如果我们相信，在同一生活的一些部分之间，可能要么有较小的补偿，要么甚至没有补偿，我们就是在变更有关补偿的主张的应用范围。鉴于还原论者的观点的内容，这是一个在正确方向上的应用范围的改变。

我们可能接着赋予这个主张较少的权重。我们的根据会是我早先提出的那个根据。补偿预设了个人同一性。根据还原论者的观点，我们相信历时的个人同一性这一事实更不深刻，或者涉及更少。因此，我们可能主张这个事实具有更少的道德重要性。鉴于这个事实是由补偿所预设的，我们会主张说补偿这个事实本身在道德上是更不重要的。尽管有关补偿的主张不能被否定，但是会如此这般被赋予更少的权重。（这里再举一个有关这一区别的例子。惩罚无辜的人是不公正的这一点不能被否定。但是可以不赋予该主张任何权重。我们不能否认这个主张这一点并不迫使我们相信该当赏罚这一点。如果我们并不

相信该当赏罚这一点，可能是因为我们是决定论者，我们可以主张，“尽管惩罚无辜者是坏的，但惩罚有罪者也一样是坏的。”)

现在转到对平衡的诘难。与补偿不同，更大的道德权重这一概念并不预设个人同一性。对平衡的诘难因而能够被否定。

这一否定可以这样表述：“我们的那些负担不能被某个他人的纯然受益而得到补偿。但是它们可能在道德上被这样的利益所超过。给予利益而不是解除负担仍然可能是对的。如果负担事实上被这些利益所超过，那么负担在道德上就被利益所超过。所需要的一切只是那些利益大于那些负担。这两者落到谁身上本身并不重要。”

这是功利主义者的回应。它可能会是他对许多论证的回应，在那些论证中对平衡的诘难与有关补偿的主张似乎不加区分。于是罗尔斯使用“不能得到正当性证明或者不能得到补偿”[21]这个短语。而佩里写道：“上百万人的幸福不知怎的完全无以补偿、甚至无以减轻一个人所受的折磨。”[22]佩里好像把这个无可否认的主张与对平衡的诘难划上了等号。这是一个错误。

还原论者的观点给予功利主义者的回应以某种支持。对平衡的诘难有几分基于有关补偿的主张。还原论者的观点既支持有关补偿的应用范围较小的主张，又支持补偿的道德权重较小的主张。补偿的范围和权重比其在倘若非还原论者的观点为真的情况下所具有的范围和权重要小。既然补偿在范围和权重这两个方面在道德上的重要性都较小，那么支持对平衡的诘难的力度就较小。我们因而可以主张功利主义者的回应比其在倘若非还原论者的观点为真的情况下似乎具有更大的可信度。但是这个主张并不蕴含着我们必须接受功利主义者的观点。此乃这个主张只是给予这个观点某种支持的原因。

这些主张可以用一个与之不同的方式加以解释。甚至那些对平衡加以诘难的人们认为，为了他生命后期自己更大的利益对一个孩子施加一些负担是可以得到合理证明的。他们的主张是，一个人的负担尽

管在道德上可以被他享受的利益所超过，但是永远不能被别人所享受的纯然利益所胜过。即便在那些人的利益远远大于那个人的负担的时候这也被认为如此。该主张于是赋予诸生命之间的边界线——或者非同一性这一事实——以超过一切的重要性。在同一生命范围内它允许，而对不同的生命它完全禁止。

照非还原论者的观点，这个主张似乎会更加令人信服。既然同一性这一事实此处被认为更加深刻，非同一性这一事实具有这样的重要性看起来似乎更加合理。照这个观点，一个人的所有的生活都同样多地是他的生活这一点是一个深刻的真理。如果这个真理——每一生命的统一性——给我们留下深刻的印象，诸生命之间的分界看来将更加深刻。它支持这样一个主张，即在道德运算中这些分界是不能逾越的。照还原论者的观点，这个真理留给我们的印象更少。我们把每一生命的统一性当作在性质上是更浅一些的，而且当作一个程度问题。因此，我们可以认为，诸生命之间的那些分界较不像，比方说，国际象棋棋盘上的方格——划分开什么是纯白的和什么是墨黑的，而更像不同国家之间的疆界线。它们于是好像在道德上具有较少的重要性。

可以如此诘难：

还原论者主张每一生命的诸部分具有较不深刻的统一性。但是他并不主张不同的生命之间具有更多的统一性。诸生命之间的诸分界，照其观点，同样地深刻。

我们可以回答如下：

如果某个统一性较少深刻，相应的非统一性亦然。我们过着不同的生活这个事实无非是我们不是同一个人这个事实。如果个人同一性这一事实是较少深刻的，非同一性这一事实亦然。这里并不存在两个不同的事实：一个事实照还原论者的观点较少深刻，而另一个事实保持同样地深刻。只存在一个事实，以及对这个事实的否定。个人的独立性就是对我们都是同一个人的否定。如果个人的同一性这个事实较少深刻，对这一事实的否定亦然。

116.对平等原则赋予较少权重的一个论证

现在转向一个不同的原则，就是在同样应得的人们之间平等分配的原则。我们绝大多数人赋予平等原则的只是某种权重。譬如，我们认为，如果不平等能够在利益总量中产生充分的获益，它是可以得到正当性证明的。

照此观点，我们并不拒斥功利主义原则。我们同意，利益总量中的任何增加都具有道德价值。但是我们坚持认为，权重也必须赋予平等原则。尽管福祉的任何获益都重要，但是谁获益也重要。某些分配，我们主张，在道德上更加令人信服。我们应该优先帮助那些并非因他们自身的过错而情形恶化的人们。而且我们应当努力追求平等。

功利主义者们会回应说："这些主张似乎令人信服。但是它们所推崇的政策正是那些倾向于增进总体福祉的政策。这种巧合表明，我们应该改变我们关于这些主张的地位的观点。[23]我们不应当把它们当作对我们的终极道德目的的限制，而是应当当作对我们的终极道德目的的指导。我们固然应该珍视平等分配。但是价值在于其典型效果。"

这个回应或许可以展开如下。我们绝大多数人相信，如果当某事发生时具有的一个纯然差异不能影响所发生的事情的本质的话，该差异在道德上不可能是重要的。对"什么时候？"这一问题的某些回答当然是重要的。我们不能忽视事件的时机性。而且甚至如此主张似乎是合理的：如果我们在筹划什么时候赋予或者得到利益，我们应当追求历时的平等分配。但是我们追求此点只是因为其效果。我们并不认为不同时间的利益的平等本身在道德上是重要的。

功利主义者们或许说："如果当某事发生的时候，它本身并不重要，那么它发生在谁身上为什么重要？这两者都是纯位置差异。重要的是所发生的事情的性质。当我们在社会政策之间进行取舍的时候，我们只需要关注利益和负担将有多么巨大。它们来自何处，是否在空间中，在时间中，抑或在人们之间，本身没有什么重要性。"

那么争执之所在部分是这样的。非功利主义者们把"谁？"的问题与"什么时候？"的问题当作十分不同的问题。如果请他们就道德上相

关的那些事实进行可能的最简单的描述，他们的描述可能是无时态的，但是它一定是有人称的。例如，他们或许说，"这个人享受的一个利益、某个别人所享受的同样利益、第一个人所承担的同等大的一个负担……"而功利主义者们或许只是说，"一个利益、同样的利益、同等大的一个负担……"

有许多支持和反对这两个立场的不同论证。我要问：成为还原论者会支持这两个立场中的一个立场吗？

我认为是会的。照还原论者的观点，把"谁"的问题比作"什么时候"的问题，而且以非人称的方式描述那些道德资料，似乎更加令人信服。这比其在倘若非还原论者的观点为真的情况下似乎更加令人信服。

再回到休谟的比较。我们绝大多数人认为，一个民族国家的存在涉及到的只不过是一系列相互联系的人的存在而已。我们并不否认诸民族国家的真实性。但是我们的确否认它们是独立地、分别地实在的。它们的存在就只涉及到以某种特定方式一道生活在其疆域中的它们的公民的存在。

这个信念支持某些道德主张。如果一个民族国家只不过是其公民的话，那么把民族国家本身当作尽职的首要对象，或者当作权利的拥有者，就不再那么似乎令人信服了。把注意力集中在公民身上，把他们较少地当作公民而较多地当作人，似乎更加令人信服。照这个观点，我们可能认为，个人的国籍在道德上有较少重要性。

照还原论者的观点，我们持有相似的一些信念。我们认为一个人的存在只是涉及相互关联的精神和物理事件的发生。我们并不否认人们存在。而且我们同意，我们不是一系列的事件——我们不是思想和行动，而是思想者和行动者。但是这之所以成立，只是因为我们通过把思想和行动归属于人们而描述我们的生命。正如我已经论证的，我们可以对我们的生活给予一个完全描述，这个描述是非人称的：并不主张个人的存在。我们否认我们并不只是从概念上不同于我们的躯

体、行动和经验，但也是独立真实的。我们否认个人是这样一个实体：他的存在独立于他的大脑、躯体以及他的经验的发生。而且我们否认个人的持续存在是一个深层的进一步事实，它必定要么全有要么全无，而且不同于物理的和心理的持续性的那些事实。

这些信念支持某些道德主张。当从道德上考虑的时候，更少地聚焦于个人——经验的主体，代之以更多地聚焦于诸经验本身，变得似乎更加令人信服。就像我们忽略人们是否来自相同的还是不同的国家是正确的一样，我们忽略诸经验是否来自相同的还是不同的生命也是正确的，此点变得似乎更加令人信服。

考虑一下受难的解除。假定我们只能够帮助两个人中的一个。如果我们帮助第一个，那么我们将更有成效；但过去受难更多的却是第二个人。那些相信平等的人们会决定帮助第二个人。这将是不太有效的；这样，两个人的生活中受难的总量将变得更大；但是每一个生活中的受难量将得以更加平等。如果我们采纳还原论者的观点，我们可能作出不同的决定。我们可能决定尽我们最大的力量来减轻受难。

为表明其原因，我们可以对这个例子改动一下。假定我们只能够帮助两个国家中的一个。那个我们最能帮助的国家是一个在已往那些世纪的历史中更为幸运的国家。我们中的绝大多数人不会认为，令人类受更多的苦，以便在不同国家的历史之间更加平等地划分苦难是正确的。在努力减轻苦难的过程中，我们并不把那些国家当作道德上重要的单位。

照还原论者的观点，我们把人们的生活比作那些国家的历史。因而关于它们我们可能抱有同样的想法。我们可能认为，当我们在努力消除苦难的时候，那些个人不是道德上重要的单位，生活也不是道德上重要的单位。我们可能再次决定追求尽可能少的苦难，不管其分配如何。

罗尔斯写道："功利主义并不在个人之间进行严格的区分。"[24]如果"个人的单独性……是道德所需的那个基本事实"[25]，那么这就是一个严重的指控。我曾尝试说明，如果功利主义者诉诸有关个人本性的真理的话，他们怎样可能提出某种辩护。

正如我已经主张的，这不可能是一种完全的辩护。问题是我们是否应当接受那些分配正义原则，而且，如果应当的话，我们应该赋予这些原则多大的权重。我主张说，照还原论者的观点，赋予这些原则较少一些权重、甚至根本不赋予什么权重似乎更加合理。但这只是意味着"照还原论者的观点本该似乎更加合理"而已。这与我有关即使照还原论者的观点不赋予这些原则什么权重也似乎不合理是相一致的。

我所提供的论证不能够表明我们必须接受功利主义者的观点。它只能够产生有关相对可信性的一些结论。当我们不再认为个人是些独立存在的实体的时候，功利主义者的观点变得更加具有可信性。似乎可信性的增益是大还是小？我的论证对这个问题不下定论。

117.一个更加极端的论证

我用另一个论证作结束。当我们不再是非还原论者的时候，我们并不只是认为个人同一性是较少深刻的，或者所涉及的较少。我们最终认为，个人同一性并不涉及深层的进一步事实。

回到有关什么为在一个生命之中取得最大化提供正当性证明的问题。施加给孩子们巨大的负担，以便他们成年的时候获得更加巨大的利益，这为什么是确当的？一些人认为这一点只能被这些人的生活的统一性给予正当性证明。当那些孩子们长大的时候，他们的负担将因这些更大的利益而得到完全补偿。

可能人们会提出一个不同的主张。或许有人主张为取得最大化提供正当性证明的是那深层的进一步事实。只有这个事实使某些人后来能够获益，以补偿先前的某种负担。补偿并不只是预设个人同一性。照这个观点，它预先设定了那深层的进一步事实[26]。

回到我出现分裂的那个事例。假定我们相信非还原论者的观点，而且假设我将是老右。在分裂之前，我拥有比我公平占有的份额要多的资源，奢侈地生活了许多年。在分裂之后，我和老左每一个将获得比公平占有的份额要少的资源。就我的情况而言，这是得到正当性证明的主张，因为它将有这样的结果，即在作为整体的我的生活中我将获得一个公平的份额。我现在较少的份额因我分裂之前预先得到的更大份额而得到完全补偿。

关于老左我们能够令人信服地持同样主张吗？心理的连续性使补偿即使在个人同一性缺失的情况下也成为可能的吗？下述回答是可辩护的：

> 不。老左从未享有一个更大的份额。他并未享受这些年的奢侈。与他能够虚拟记得你享受这种奢侈这一点并不相关。他与某个在老左并不存在的时候享有比他的公平份额更多的人在物理上和心理上保持连续性这一点并不相关。现在给予老左一个比公平的份额要少的份额会是不公平的。在缺乏个人同一性的情况下，心理的连续性并不能够使补偿成为可能。

下面假设我们最终认可了还原论者的观点。我们业已可辩护地主张说，只有那深层的进一步事实使补偿成为可能。我们业已主张过，就像老左的事例所表明的，物理的和心理的连续性本身并不能够使补偿成为可能。我们现在认为没有什么深层的进一步事实，而个人的同一性就寓于这两种连续性之中。既然我们能够可辩护地主张只有这个深层的进一步事实使补偿成为可能，而这样的事实并没有，那么我们能够可辩护地得出结论说，不可能有历时的补偿。我们能够主张，某个时间的利益不能为另一个时间的某个负担提供补偿，即便两者出现在同一个生命之中。只可能有共时的补偿，就像当我暴露在刺骨的寒风中的脸庞的疼痛被我所看见的攀登过的山脉的壮丽景观所完全补偿那样。[27]

既然这个结论是可辩护的，这个论证就比能够支持功利主义者的观点的那些论证有力得多。那些论证并未表明这个观点是可辩护的。

但是，就像早先诉诸那深层的进一步事实的那些论证的情况一样，尽管这个新结论是可辩护的，它也能够是雄辩地予以否定的。

这个新结论涉及到我们的那些分配原则的又一个范围变化。这些原则的应用有赖于可能的补偿的范围。照刚刚给出的论证，即便在同一个生命中也不可能有历时的补偿。考虑一下平等原则。照刚刚给出的论证，我们既不应当追求不同的人的生活中的平等，也不应当追求同一人的生活中不同部分之间的平等。我们应当追求人们在特定时间所处的那些状态的平等。

内格尔作了一些相关的评述：

> 要注意，这些思想并不依赖任何历时的个人同一性观念，尽管它们能够利用这样的一个观念。要引发它们，所需要的只不过是某个时间诸个个人之间的区分。只要我们能够区分出被两个个人所拥有的两个经验与被一个个人所拥有的两个经验，分配平等的冲动就会产生。历时的个人同一性的准则只是决定某个分配原则进行操作的单位尺度的大小。简言之，这就是我所认为的帕菲特在有关分配正义和个人同一性之间关系的说明方面的错误之处。[28]

在内格尔所讨论的那个说明中，我区分了变更我们的道德原则的应用范围与赋予这些原则以一个不同的权重。而且我主张说，如果我们从非还原论者转到还原论者的观点，我们能够更加令人信服地赋予这些原则以更大的范围，但是更小的权重。[29]内格尔主张，所产生的效果只是赋予这些原则更大的范围——我主张某个观点的改变也可能影响到我们赋予这些原则的权重是错误的。但是，我的主张有什么错？所产生的效果为什么应当只及于范围？一个有关那些事实的观点的变化经常使赋予某个道德原则一个不同的权重这一点是可信的。在现在这个事例中，如果这不能令人信服，就需要予以证明。我认为这是不会得到证明的。

我现在讨论的论证支持内格尔所认同的那种效果。它诉诸有关个人同一性准则的一个观点的变化。通过变成还原论者，我们不再相信个人同一性涉及那深层的进一步事实。该论证主张，补偿所预先设定

的不是照任何一种观点而言的个人同一性，而是照非还原论者的观点而言的个人同一性。补偿以深层的进一步事实为预先设定。心理连续性在这个事实缺失的情况下不能够使历时补偿成为可能。

如果这是不可能的，那么我们的那些分配原则将告诉我们去做什么？它们大致上与否定的功利主义——即赋予解除苦难以优先权的观点相吻合。内格尔谈到了分配原则操作的单位。如果这个单位是一个人的生活的整体，就像罗尔斯和其他许多人所假定的那样，那么平等原则将告诉我们要去努力帮助那些境况越来越差的人们。如果所说的单位是任何一个人在某个特定时间下的状态，那么，平等原则将告诉我们要去努力改善的不是那些日渐糟糕的人们的生活，而是人们所处的最糟糕的状态。

平均主义者可能不同意我们应当努力使人们平等的那个方面。我们应当追求福祉的平等，还是资源的平等？假设我们一度认为分配原则的相关单位是一个一个完整的生命，而现在我们认为相关的单位是人们在特定时间所处的状态。我们这种观点上的改变使主张我们应当关注资源的平等变得可信度较少。某个占有大量资源的人在特定的时间的状态可能相当糟糕，他的那些大量资源可能对缓解这种状态无能为力。如果相关的单位是人们在特定时间所处的状态，那么主张我们应当关注在这些时间内人们经验的质量则可信度更大。与这些境况渐糟的人们相呼应的，照罗尔斯学派的观点，将是那些最糟糕或者最不可取的经验。这些经验要么将是身体上的剧痛，要么将是精神上的极度苦闷或悲伤。我们的平等原则将告诉我们要去努力避免或者改善这些经验，而且较之促进可取的经验要给予这一点以优先权。

当我们把单位从一个一个的整体生命变为人们在每一个特定时间的经验的时候，我们应当赋予我们的那些分配原则以一种不同的权重吗？至少有一位学者主张我们应当这么做。瓦克斯贝格主张，如果我们以这种方式变更我们的分配原则的范围，我们就不应当赋予这些原则任何权重。如果相关的单位是人们在任何一个特定时间下的经验，

瓦克斯贝格认为，那些分配原则便失去了其令人信服之处。[30]我更加谨慎地认为，这些原则变得较不可取。在我看来，解除苦难应当具有比普通的或曰肯定的功利主义者所赋予的权重以更大的权重是令人信服的。但是，它应当具有绝对的优先权、甚至具有更进一步的权重则看来并不令人信服。

哈克萨尔提出了照还原论者的观点分配原则应当具有更少权重的一个根据。他写道：

> ……如果帕菲特的……理论是正确的，如果没有那些持续性的一个一个的个人（除了在微不足道的意义上），我们为什么对世界上的苦难如此愤愤不平？苦难仍然会是真实的，但是当（内在地）那同一个人不断地受难的时候又更糟多少啊！[31]

照我现在所讨论的论证，在没有那深层的进一步事实的情况下，不可能有历时的补偿。既然没有这样的事实，就没有这样的补偿。如果我们如此主张，我们就不能同时主张：如果一定数量的苦难集中在某个特定的个人生活中并且延续下去而不是分散在不同生命中的时候，这一定量的苦难将是一种更大的恶。有关这是一种更大的恶的主张假定相关的单位是历时地持续的，而这正是上述论证所否定的。如果苦难的糟糕性只是在某个特定时间所具有苦难的糟糕性的话，哈克萨尔的主张看起来令人信服。照此观点，尽管集中的和延续的苦难是糟糕的，但是并不像倘若非还原论者的观点为真的时候那样糟糕。这减少了否定的功利主义的可信性。

这些评述容易遭到误解。照无论是关于个人同一性的还是关于分配原则的范围的任何观点，如果某个人知道苦难将延续下去，这个人的苦难事实上将更加巨大，更加难以忍受。这为人们优先努力避免这样的延续受难提供了一个强有力的证据，即使对肯定的功利主义者而言也是如此。我的评述一直是关于我们有没有进一步的道德理由来优先避免苦难，无论苦难是否延续下去。我们具有这样的理据正是否定的功利主义者所主张的。我们业已发现，如果我们变更我们平等原则

的范围，从而相关的单位变成人们在特定时间所处的状态，我们的平等原则大致与否定的功利性原则相契合。我们应当赋予这两个原则什么权重？与瓦克斯贝格不同，我认为这些原则仍然令人信服。但是我认为它们具有的令人信服之处比当我们认为那相关单位是一个人生活的整体的时候它们会有的令人信服之处要小。而且这有几分是出于哈克萨尔的理据。如果一个人生活的统一性并不涉及那深层的进一步事实，有关不可能有历时的补偿的主张就是可辩护的。而且，如果这种统一性不能使补偿成为可能，它也不可能使倘若非还原论者的观点为真的情况下延续下去的苦难会形成的完全的恶成为可能。

118.结论

我现在将对这个颇长讨论的一些主要主张总结一下。我问了有关个人同一性的观点的变化如何会影响到我们的情绪和我们有关合理性和道德性的信念。而且我刚刚讨论了，如果我们改变我们的观点的话，这如何影响到我们有关平等原则和其他一些分配原则的信念。我稍早论证过，如果我们从非还原论者的观点转到还原论者的观点，那么主张在同一生命中补偿的范围较小则变得更令人信服。于是，有关施加在某个孩子身上的巨大负担不能够由这个孩子成年生活中的更大一些的利益补偿或者不能够完全补偿的主张更加令人信服。当我们如此扩展那些分配原则，以便它们既涵盖一个个的整体生命又涵盖同一生命中联系微弱的诸部分的时候，这使这些原则更为重要。这是脱离功利主义观点的一个步骤。

我刚刚讨论了改变我们的分配原则范围的第二个论证。其主张是，只有那深层的进一步事实才使跨越某个生命的不同部分的补偿成为可能。既然没有这样的事实，我们应当像内格尔所建议的那样，改变"某个分配原则进行操作的单位尺度的大小"。那些单位缩小到人们在特定时间的状态。这个结论是可辩护的，但是对其否定亦然。

鉴于在其范围中的这一变化，平等原则大致与否定的功利主义相契合。如果我们把平等原则运用于人们在特定时间的状态，我们应给予该原则更多权重吗？我提议不要。

第三个论证主张，不论分配原则的范围如何，我们都应给它们较少的权重。这些原则通常被认为基于不同的个人之间的独立性或曰非同一性。照还原论者的观点而言这个事实较不深刻，因为同一性较不深刻。它并不涉及我们倾向于认为的那深层的进一步事实。既然它们所基于的事实被看作是较不深刻的，赋予那些分配原则较少的权重就更加令人信服。如果我们不再认为个人是独立存在的实体，而且最终认为某个生命的统一性所涉及的不过是这个生命中不同的经验之间的各种各样的关系，那么更多地关注经验的质量、更少地关注谁的经验就更加令人信服。这给予功利主义的观点某种支持，使之比倘若非还原论者的观点为真时更加令人信服。即使我们没有意识到这一点，我们当中的绝大多数人也倾向于认为非还原论者的观点。功利主义的非人称性因而比我们绝大多数人所信以为是的更少令人信服。

稍早我讨论了如果我们改变我们的观点的话，我们是否具有同样多的理据来特别关心我们自己的未来。一些学者主张说，只有那深层的进一步事实才给予这种特别关心以正当性证明，而且，既然没有什么这样的事实，我们没有什么理据来特别关心我们自己的未来。这一极端主张是可辩护的，但是对其否定亦然。

我然后提出了针对有关合理性的自利论的另一个论证。所诉诸的是这样一个事实，即个人同一性中的重要部分，心理联系性，随着时间的过去保持在一些减弱的程度上。当某个重要事实保持在一个减弱的程度上的时候，认为这一事实具有较少的重要性不可能是非理性的。因而，现在较少关心我们的未来中那些与现在联系较不密切的部分不可能是非理性的。鉴于还原论者的观点的真理性，这驳倒了古典自利论。

照修正的自利论，这是没有被驳倒的，做自知将对自己更糟的事可能并非是非理性的。极大的轻率可能并非是非理性的。如果这样的一些行动并非非理性的话，它们需要进行评判。我提出来的主张是，我们应当把它们当作在道德上是错误的。如果这样的行动在道德上是错误的，这强化了那个家长制的事例。

如果我们成为还原论者，我们能够令人信服地主张受精卵不是人类的一员，只是在怀孕期间逐渐变成人。这支持有关在怀孕的前几周堕胎并无不当、只是逐渐变得不当而已的主张。

我还考察了照还原论者的观点有关该当赏罚和承诺我们应当相信什么。一些学者主张，只有那深层的进一步事实带有该当赏罚，而且既然没有这样的事实，我们不能够为过去的罪过而该当受罚。这个结论看起来可以辩护，但是对其否定亦然。然后我论证了那个一般性的主张，即如果在罪犯现在与他犯罪时候的自我之间的联系是较弱的，他该当较少地受惩罚。相似的主张也适用于承诺。

我们应当是还原论者。如果这是一个观点的变化，那么它支持我们有关合理性和道德性的信念中的数个变化。还有其他一些变化我尚没有进行讨论。[32]

对我们的情绪所产生的效果可能因人而异。对于那些接受极端主张的人们而言，效果可能是令人担忧的。我描述了对我的效果。既然我否定这些主张，我发现那个真理是解放性的和抚慰性的。它使我更少关心我自己的未来和我的死亡问题，而更多地关心他人。对这种我所关心的方面的拓展，我表示欢迎。

---

[1]见Perry（1），第102页。

[2]见Perry（1），第112页。

[3]Madell，第116页。

[4]Haksar，第111页。

[5]方济各会（St. Francis），是天主教托钵修会之一。在教皇批准下，由意大利人方济各创立于1209年。提倡过一种安贫乐道、禁欲苦行的生活。因会士之间互称小兄弟，故又得名“小兄弟会”。——译者

[6]见其《神与灵》，第4页。

[7]奥德修斯（Odysseus），希腊神话人物，荷马诗史《奥德赛》中的主人公，特洛伊战争中希腊大军的领袖之一。在希腊人对特洛伊城久攻不下的情况下，一日献木马计，城池得破，希腊人获胜。——译者

[8]塞壬（Siren），希腊神话中半人半鸟的海妖。常用令人难以抵挡的美妙歌喉诱惑过往的航海者去一睹芳容，以使航船触礁沉没。荷马诗史《奥德赛》中的主人公奥德修斯渡海远征特洛伊途中，为避免经不住她令人难以抵挡的歌喉诱惑而导致触礁沉船，命人把自己绑在桅杆上。——译者

[9]Nabokov，第64页："他们说这个英国人在世界上所爱的惟一东西是俄罗斯。很多人不明白他为什么不住在那里。穆恩对此类问题的回应总是千篇一律。'问问罗宾逊'（那位东方学学者）'为什么不呆在巴比伦。'被提出的完全合情合理的诘难会是巴比伦不再存在。穆恩会带着狡黠的、不露声色的微笑点点头。他在布尔什维克暴动中看到了某种清晰的终结性。尽管他情愿允许某种文明在那些原始阶段之后将来可能会在'苏维埃联盟'中发展出来，但仍然认为俄罗斯已经完结，而且是不可重复的。"

[10]Nozick（2），第33页写道："我主张，在道德上对于我们可以做什么的那些边-约束（side-constraints），反映出我们的独立存在这个事实。它们反映出这样一个事实，就是说，在我们中间没有什么在道德上起平衡作用的行动能够发生；在道德上没有什么他人如此重于我们中的一个人的生命，以至于导致一个更大的全面的社会的善。我们中的一些人为他人所作的牺牲没有什么是得到正当证明的。这个草根观念就是说，存在着具有独立生命的不同的个人，因此没有什么人可以为他人作出牺牲……"（第二个强调为我所加）。

[11]Rawls，第27页。

[12]Gauthier（2），第126页。

[13]Espinas，引自Perry（3），第402页。

[14]Gauthier（2），第126页。

[15]Rawls，第28及141页。

[16]Rawls，第27页。

[17]Nagel（1），第134页。

[18]Rawls，第190页。

[19]Rawls，第29页。

[20]Rawls，第28页。我省略了"宛如他们是一个人"这句话，因为我是在问这个推理是否必须包括这个假定。

[21]Rawls，第61页。

[22]Perry（3），第671页。

[23]参见Sidgwick（1），第425页，"除非我们把源于功利主义和常识相契合这一复杂特征的复证力量充分考虑进去，否则功利主义的论证就不能够得到公正判断。"

[24]Rawls，第27页。我对另一个说法添加了评论。罗尔斯写道："存在……一个令人好奇的反常状态……习惯上功利主义被认为是个体主义的，这样想当然有些不错的理由。功利主义者们……持有的观点是，社会的善是由诸个体所享受到的利益构成的。然而功利主义并不是个体主义的……因为……它把对于一个人而言的选择原则应用于社会"。（第29页）

我的说明涉及这一解释。个体主义者主张（1）社会的福祉只寓于其成员的福祉之中，而且（2）那些成员具有享受公平份额的权利。假设在关于社会和国家方面我们是非还原论者。我们认为一个社会或者国家的存在超越其成员的存在。这个信念威胁到主张（1）。它支持这样一个观点，就是说，对于该社会或者该国家而言是善的东西超越对于其成员而言是善的东西。这又威胁到主张（2）；在对于一个超越的国家目标的追求之中，公平份额可能看起来不再那么重要。在国家方面的非还原论者可能因而对个体主义者的两个主张都加以拒斥。功利主义者拒斥（2），但是采纳（1）。如果功利主义者的观点依赖于关于国家的非还原论的话，正如罗尔斯所主张的，这会是"一个令人好奇的反常状态"。如果这就是他们的根据，那么我们应当期待他们对这两个主张都予以拒斥。我已经描述过一个不同的根据。功利主义者不是关于国家的非还原论者，而可能是关于人们的还原论者。这会消除那个反常状态。功利主义者也是关于国家的还原论者，而这个双重的还原论给予功利主义者的观点某种支持。如果我们是关于国家的还原论者，那么我们能够更加令人信服地接受个体主义者的第一个主张：一个国家的福祉只寓于其公民的福祉之中。如果我们还是关于人们的还原论者，那么我们能够更加令人信服地拒斥第二个主张，即对公平份额的要求。我们可能把较少一些的权重赋予各自生活的统一性，以及不同生活之间的差异，而把更多的权重赋予那些五花八门的经验，这些经验一道构成这些生活。于是我们可能断定，道德上重要的只是所发生的事情的性质，而不是发生在谁的身上。我们可能裁定，增加利益减少负担总是对的，而无论其分配如何。参较安斯舒茨这样一种说法："边沁的个人主义原则"，与穆勒的不同，"完全是过渡性的"，因为"边沁是在说……正如一个共同体可以还原为那些据说是其成员的那些个体，至少出于道德和立法的目的，个体也可以还原为他们据说享受到的那些快乐和遭受到的那些痛苦"。（第19—20页）边沁的功利主义可能部分地得到他对还原论者的观点所抱信念的支持。但是我的主张不能适用于所有的功利主义者。西季威克就是一个明显的例外。在《伦理学方法》第416—417页上，西季威克赋予平等原则以某些权重。但是功利准则具有绝对的优先权。西季威克对分配原则的拒斥不能够以我所讨论的那种方式加以解释。西季威克拒斥休谟的还原论者的观点。西季威克这本著作的第一版结尾之所以用"失败"一词，主要是因为他分配给人们之间的区分以这样的权重。（参见Sidgwick（1），第404页，或者第498页有关"任何一个人与另一个人之间的区分都是真实的和根本的"这一说法。）正如我已经说过的，个人之间的分离性在西季威克看来足以深刻地支持自利论者对那些道德主张的拒斥。我宁愿我们会问他为什么这个深刻的事实在他看来并不支持那些公正分配主张。

[25]Findlay，第294页。

[26]瓦克斯贝格为这个观点辩护。瓦克斯贝格添加并且令人信服地辩护这样一个主张，就是，对于那个进一步的事实的信念涉及到所有的经验都是发生在"同一意识"之中这样一个观念，而该意识历时的同一性是对意识在一个时刻的统一性的一种误概括。

[27]L·特姆金（Temkin）（在书信往来中）论证说，如果一个还原论者否认历时的补偿，他也应当否认共时的补偿。意识在一个时刻的统一性并不是一个比历时的心理连续性更加深刻的事实。当我们考虑到对最近的那些时刻或者似是而非的当前（the specious present）的短期记忆的时候，这一点就看起来得到了说明。在这个短期连续性和意识的一个时刻的统一性之间不能有一种深刻的差异。这支持下面那个提法，就是，按照这个观点我们

的分配原则大致与否定的功利主义相契合。如果能够有共时的补偿，就像在我所说的那个刺骨的寒风和壮观的景色的事例当中一样，最糟的那组经验在一个单一的时刻无需是那些包含最强烈的痛苦或受难的经验，因为这个痛苦或大于这组经验中其他一些经验所补偿的。如果甚至不能够有共时的补偿，我们所关心的不会是去使最糟的那组经验更好，而是去使这样的一些组别中的最糟的成员更好；而且这些更有可能是强烈的痛苦，或者其他种类的极端痛苦。

[28]Nagel（4），第124—125页，脚注16。

[29]Parfit（3），第153页。

[30]Parfit（3），第153页。

[31]Haksar，第111页。

[32]对于那些并非关于合理性和道德的信念的影响，我尚未讨论。这样的一个影响可能是减少他人之心难题（Problem of Other Minds）。当我是一个非还原论者的时候，我认为我的生活的统一性并不只是寓于我的经验与我的大脑当中的诸状态和过程之间各种各样的联系。我认为，所有的这些经验是我的经验这一点是一个进一步的、根本的事实。这使我更加难以想象某些经验不是我的经验会是什么，这给他人之心难题添加了权重。按照还原论者的观点，这个困难则消失了。对于他人的过去经验的类记忆之可能性还有可能帮助化解这个难题，就像我出现分裂的那个设想的事例一样。结果产生的两个人中的任何一个是否会是我这个问题是一个空洞的问题。（正如我所解释的，即便一个答案是最好的描述，这也是一个空洞的问题。）既然这些人们的明显记忆是否纯粹是对另一个头脑的内容的类记忆会是一个空洞的问题，这些人们就不会严肃地看待他人之心难题。


第四编 未来的世世代代
上海译文出版社

第十六章

非同一性问题

还有一个有关个人同一性的问题。我们中的每一个人可能就从未存在过。使之为真的会是什么呢？这个问题的答案产生了被我们绝大多数人所忽略了的一个问题。

第四编的目的之一就是讨论这个问题。另一个目的是讨论这个问题由之产生的那个道德理论部分。这是涵盖我们如何影响未来的世世代代的部分。这是我们的道德理论中最为重要的部分，因为未来的几个世纪将是人类历史中最为重要的世纪。

119.我们的同一性何以事实上依赖于我们何时成胎

什么本会使某个特定的人从未存在过这一点为真？带有一个限定，我相信

> 依赖时间的主张：如果任何一个特定的人当他事实上成胎的时候还没有成胎的话，他本永远不会存在这一点事实上是真的。

这个主张并非显而易见地为真。于是一个妇女写道：

> 推测我们的父母如果与他人结婚的话我们本会是谁总是令人迷惑的。[1]

在疑惑她本会是谁的时候，这位妇女忽视了这样一个答案："谁也不是。"

尽管依赖时间的主张并非显而易见地为真，但并无争议，易于相信。从而它不像有关历时的个人同一性的还原论者的观点。后者是数个竞相争胜的观点中的一个，而且难以令人置信。依赖时间的主张不是关于历时的个人同一性。它涉及的是一个尽管相关但却不同的主

题：世界的各种不同的可能历史中的个人的同一性。关于这个主题有多个观点值得讨论。但是依赖时间的主张并非是这些观点中的一个。它是一个就所有这些观点而言都为真的主张。

如同我已经说过的，该主张应当予以限定。我们每一个人都成长于一对特定的细胞：一个卵子和来自数以百万计的精子中的那个使之受精的精子。假设当我的母亲在事实上孕育了我的时候并没有孕育一个孩子。而且假设她这次数天之内孕育了一个孩子。这个孩子本会从我由之成长的那同一个特定卵子成长起来。但是，这个孩子的成胎即便早了或者迟上数秒钟，也几乎可以肯定，他本会从一个不同的精子成长起来。这个孩子本会具有我的一些基因，但并非所有的基因。这个孩子本会是我吗？

我们倾向于认为任何关于我们的同一性的问题都一定有一个答案，这个答案必定要么是要么否。一如既往，我拒斥这个观点。在其中存在着我们的同一性并不确定的一些事例。我刚刚所表述的就可能是这样的事例。如果是的话，我的问题就没有答案。如果这些事件发生了的话，我本永远不会存在——这既不真也不假。虽然我总能问"我本会存在吗？"但在这里却是一个空洞的问题。

最后这些主张是有争议的。既然我想让依赖时间主张无争议，我将搁置这些事例。该主张可以变成

（TD2）如果任何特定的人在其事实上成胎的1个月的时间内尚未成胎的话，那么他在事实上本永远不会成胎的。

我认为，这在事实上是真的。我并非认为它必然地为真。关于这个主题的不同观点形成有关必然是什么的一些相互竞争的主张。正是因为我主张的更少了一些，所以我的主张才不带有争议性。在有关什么本来能发生的问题上那些意见不一的人们会在有关什么本来会事实上发生的问题上看法一致。正如我将讨论的，持有所有那些令人信服的观点的人们会赞同我的论点。

这些观点形成那些有关每一个特定的人的必然属性的主张。一个人的必然属性中的某些属性为每一个人所具有：这些就是作为一个人所必须的那些属性。我们在这里所关注的是每一个特定的人的与众不同的必然属性。假设我主张P是康德的那些与众不同的必然属性中的一个。这意味着康德本不能缺乏P，而且只有康德本会具有P。

根据

起源的观点，每一个人都有这个与众不同的必然属性：即已经从这个人事实上由之生长出的一对特定的细胞生长出来。[2]

这个属性不能是完全地与众不同的。任何一对同卵双胞胎两者都生长自这样的一对细胞。而且任何一个受精卵后来也许会分裂，而产生双胞胎。起源的观点必须予以修正，以应对这个问题。但是我无需讨论这个修正。就我的那些目的而言，即照此观点康德本来就不能生长自一对不同的细胞，它已足够了。因为可能存在双胞胎，所以只有康德本会生长自这对细胞为假——这一点是不相关的。

起源的观点的持有者们会接受我的主张，即如果康德在其成胎时的1个月的时间内本没有成胎，他本会事实上从未存在过。如果他并未在那个月被孕育，就没有什么孩子事实上会从他由之生长的那对特定的细胞中生长出来。（这个主张形成了一个关于这对细胞的与众不同的必然属性和关于人类的生殖系统这两方面的假定。但是这些假定并非是带有争议性的。）

根据某些其他观点，康德本来能生长自一对不同的细胞。按照

无特征的笛卡儿式的观点，康德是一个特定的、没有什么与众不同的必然属性的笛卡儿式的自我。

照此观点，一个人的同一性与他的物理的和精神的特征没有什么联系。康德本来会是我，反之亦然，尽管，如果这发生了的话，没有哪一个人会注意到什么差别。像我所做的那样，主张我们应当拒斥笛卡儿式观点的这个版本，最糟不过带有温和的争议性而已。

其他两个观点密切相关。按照

> 描述性观点，每一个人都有多个与众不同的必然属性。这些是这个人的最重要的与众不同的属性，而且它们并不包括已经生长自一对特定的细胞这个属性在内。

在康德的事例中，这些属性会包括他对某些著作拥有著作权。这个观点的一个版本并不主张康德必定拥有所有这些属性。任何带有这些属性中的绝大多数属性的人本都会是康德。

按照

> 描述性姓名观点，每一个人的姓名都意味着"那个……的人"。现在对我们而言，"康德"意味着"那个写了《纯粹理性批判》等等的人"。某个特定的人的必然属性是当我们解释这个人名字的意义的时候会列举的那些属性。

这个观点和那个描述性观点可能与笛卡儿主义的其他版本结合在一起。康德可能被主张为笛卡儿式的自我，其与众不同的必然属性包括某些著作的著作权在内。但是这两个描述性观点无需添加这个主张。[3]

对这些描述性观点的一个诘难是，每一个人的生活本来会是非常不同的。康德本会在摇篮中夭折。既然这是可能的，某些著作的著作权就不可能是康德的必然属性中的一个属性。

对这个诘难的一个回应退却到一个更弱些的主张。可能会如此说：

> 尽管这个属性不是必然的，但却是与众不同的。康德可能本没有写这些著作。但是，在任何其中有某个单一的人写了这些著作的可能的历史中，这个人本会是康德。

我无需讨论这个回应或者其他哪一个回应反驳了这一诘难。即便该诘难能够得到答复，我的依赖时间的主张也是真的。

按照这两个描述性观点，康德本来能够生长自一对不同的细胞，甚至有一对不同的父母。如果康德的母亲在怀上康德的时候本来没有怀上任何孩子，而另一对夫妻孕育了一个后来写了《纯粹理性批判》等著作的孩子的话，这本来是会发生的。按照这些描述性观点，这个孩子本会是康德。他也许本不被称作康德。但是这并不令那些持有这些观点的人们发愁。他们会主张说，如果这发生了的话，康德本会既有一对不同的父母，又会有一个不同的名字。

尽管他们认为这或许本会发生，但是绝大多数持有这些描述性观点的人们会接受我的这样一个主张，即它事实上本不会发生。如果他们主张它本会发生的话，他们必定接受托尔斯泰在《战争与和平》的跋中所陈述的观点，即历史并不依赖于特定的人们所作出的那些决定。照此观点，如果拿破仑的母亲没有生他的话，历史本会提供一个“拿破仑的替身”，他会在1812年入侵俄罗斯。而且，如果康德的母亲没有生他的话，历史本会提供另一个写了《纯粹理性批判》的作者。这一观点太不令人信服，不值得讨论。

持有这些描述性观点的人们可能还有另一种拒斥我的主张的方式。他们可能认为，康德的那些必然属性的与众不同要少得多。例如，它们或许只是：身为他母亲的头一个孩子。这个主张答复了每一个人的生活本会非常不同的诘难。但这个主张也是太不令人信服，不值得讨论。我是我母亲的三个孩子中的第二个。这个主张蕴含着这样一种荒谬性，即如果当我的母亲事实上怀上我的时候没有怀上任何孩子的话，我本来可能是我的妹妹。

下面考虑一下那些描述性观点在其中看来最令人信服的那一种可能的历史。假设当康德的母亲孕育康德的时候没有怀上孩子，而是1个月以后她怀上了一个确切地相似康德的孩子。这个孩子本会生长自一对不同的细胞；但是出于一种令人惊讶的巧合，一种事实上永远不

会发生的巧合，这个孩子具有康德的基因中的所有的基因。而且假设，除了出生晚这一事实和产生的结果之外，这个孩子会过着就像康德所过的生活一样的生活，写了《纯粹理性批判》等等。

按照那些描述性观点，这个孩子本是康德。持有起源的观点的人可能会反驳说：

> 康德是一个特定的人。在你们想象出的历史中，你没有表明你所指称的是这个特定的人。在这个想象出的历史中，本会有某个确切相似这个人的人。但是，就像任何两个确切相似的事物所表明的那样，确切的相似性并不就是数字性的同一性。

这些评述解释了为什么起源的观点涉及一个人由之生长而来的那一对特定的细胞。

第五种观点也是这样直接指称。按照

> 反向变量观，这个指涉无需到起源点，或者一个人由之生长而来的那些细胞。指涉可以是在这个人的生命中的任何时刻。通过这样的一个指涉，我们能够描述这个人何以本会有一种不同的起源。

考虑一下一个持有这一观点的人，这个人在1780年正在上康德的一堂课。这个人可能声称：

> 康德就是正站在那里的那个人。康德本可能有不同的父母，而且直到最近的过去本会还过着一种不同的生活。想要使这成为所发生了的，所需要的只是这个不同的生活本会导致康德现在站在那里。

这个观点必然会得出一些进一步的主张。但是它回答了有关要为一个同一性主张提供正当性的证明我们需要的不止是相似性这个诘难。那些持有起源的观点的人们因而需要一种对反向变量观的不同诘难。就我所要达到的那些目的而言，我无需在这些观点之间作出决断。

按照反向变量观，康德本或许有一个不同的起源。但是这一观点的持有者们会接受我的这样一种主张，即，事实上这本就不会发生。他们会同意，如果康德在其成胎时的1个月的时间之内没有被孕育的话，他事实上本永远不会存在。

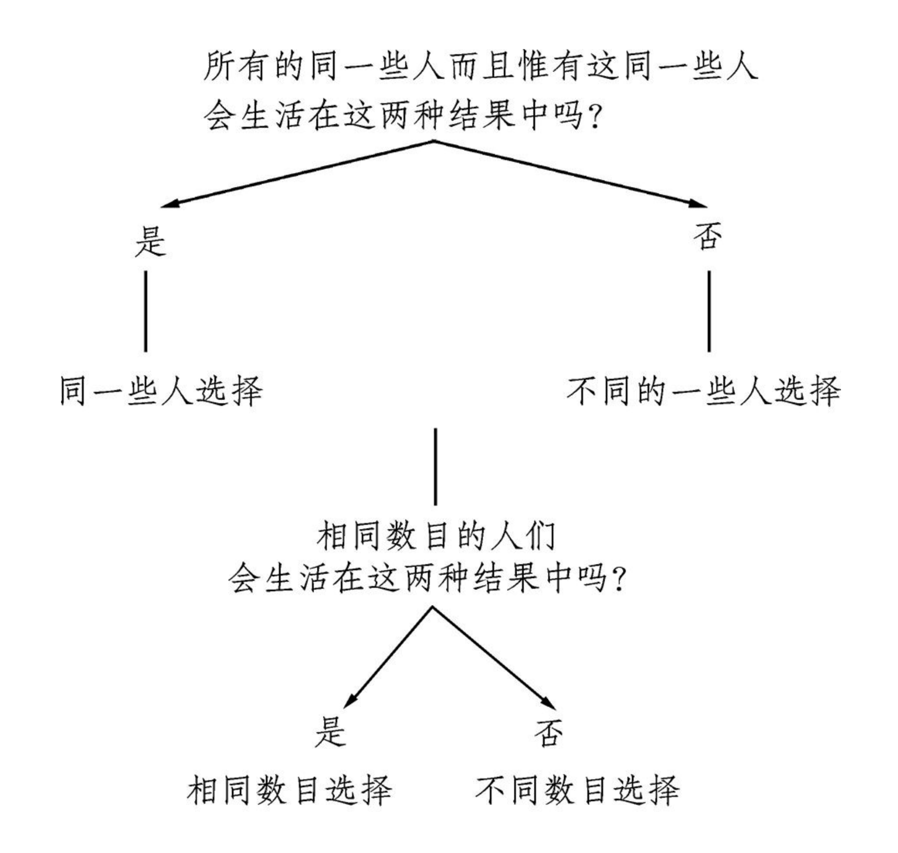
我现已表述了所有关于我们在不同的可能历史中的同一性的观点。[4]我在注释[5]中对这些观点如何与各种有关我们的历时同一性的不同观点联系在一道进行了讨论。按照所有令人可信的观点，我的依赖时间的主张是成立的。这个主张适用于任何人。你是在某个时刻成胎的。事实上的确，如果你在那个时刻的1个月的时间之内没有成胎的话，你本永远不会存在。

120.三种选择

除非我们或者某种全球性的灾难毁掉人类，否则将会有一些生活在以后的人，这些人现在并不存在。这些是未来的人。科学已经赋予我们这个世代的人影响这些人的能力，也赋予了我们预测这些影响的能力。

有两类影响引发出一些令人困惑的问题。我们能够影响到未来人的同一性，或者说，影响到谁是那些将会生活在以后的人。而且我们能够影响到未来人的数目。这些后果赋予我们各种不同的选择。

在比较任意两个行动的时候，我们能够发问：



不同数目选择既影响到未来人的数目又影响到未来人的同一性。相同数目选择影响到未来人的同一性，但是并不影响到他们的数目。同一些人选择对两者都不产生影响。

121.我们应当赋予未来人的利益什么样的权重？

我们的绝大多数道德思维都与同一些人选择相关。正如我将要论证的，这样的一些选择并不像我们绝大多数人所设想的那样数目众多。我们的许多选择事实上既将影响到未来人的同一性，又影响到他们的数目。但是在这些事例的绝大多数情况下，因为我们不能预测那些特定的影响会是什么，这些影响在道德上可能遭到忽视。我们可以把这些事例当作它们宛如同一些人选择那样来对待。

在一些事例中，我们能够预测某个行动要么可能、要么将会违背未来人的利益。当我们作出一个同一些人选择的时候，可能就是如

此。在这样的一个事例中，无论我们选择什么，所有的同一些人而且惟有这同一些人将会生活。这些人中的一些人将是未来人。既然无论我们作何选择，这些人将会存在，那么我们就能以一种相当直截了当的方式伤害这些人，或以一种相当直截了当的方式使这些人受益。

假设我在一个森林的林下植物中丢弃了一些碎玻璃。100年后，这些玻璃划伤了一个孩子。我的行动伤害到了这个孩子。如果我安全地埋掉了这些玻璃的话，这个孩子本会毫发无损地穿过这个森林。

我所伤害到的那个孩子现在还不存在这一点会造成道德上的一种差异吗?

照一种观点的说法，道德原则只涵盖那些能够往复互动（reciprocate）的人们，或者说，只涵盖那些能够相互伤害和相互受惠的人们。就像我们能够令人信服地假设的那样，如果我不可能被这个孩子伤害或者不可能得益于这个孩子，那么我对这个孩子造成的伤害不具有道德上的重要性。我假定我们应当拒斥这个观点。[6]

一些学者认为，尽管我们应当关注对这些未来人的那些影响，但是我们更少关注对更远的未来的人的那些影响在道德上是得到正当性证明的。这是福利经济中和成本-效益分析中的一个普通观点。照此观点，我们可以按照每年n%的比例贴现我们行动和政策的更加遥远的影响。这被称作社会贴现率（social discount rate）。

假设我们正在考虑如何安全地处理被称作核废料的放射性物质。如果我们相信社会贴现率，那么我们将关注的只是更近的未来的安全。我们将不会因某些核废料的放射性将长达数千年这一事实而烦恼。按照5%的贴现率，对明年的一个人的死亡的关注要大于今后500年里死亡10亿人的关注。照此观点，更远的未来中的灾变现在可以被当作在道德上是微不足道的。

就像这个事例所提示的，社会贴现率难以站住脚。时间上的遥远性大致上与诸如可预测性等重要因素相关联。但是，正如我在本书“附

录”F中所论证的，这些关联太粗糙，不能为社会贴现率提供正当性证明。未来事件在目前的道德重要性并非以每年n%的比例递减。时间上的遥远性本身并不比空间上的遥远性更为重要。假设我向森林的遥远深处射出一枝箭，伤到了那里的某个人。如果我本该知道在这个森林中或许有人的话，我为这严重的疏忽而感到负罪。因为这个人太远，我不能辨明我所伤害的那个人。但这绝不是理由。这个人离得太远也不是什么理由。有关对于那些时间上遥远的人们的影响，我们应当持同样的主张。

122.女孩的孩子

未来的人在一方面不同于距离遥远的人。我们能够影响他们的同一性。而且我们的许多行动具有这种影响。

这个事实产生了一个问题。在我描述这个问题之前，我将重述一些事先的评述。我假定一个人可以或多或少地在具有道德重要性的意义上比另一个人更糟。但是我并不认为这些对比即便在原则上是精确的。我假定，只有大致的或者部分的可比性。照此假定，两个人的确可能哪一个都不比另一个更糟，但是这并不会蕴含着这些人的处境确切地同等好。

“变得更糟”可以用来指称某人的幸福水准，或用来更为狭义地指称他的生活标准，抑或用来更广义地指称他的生活质量。既然“生活质量”是最广义的一个短语，我将时常予以接受。我也称某些生活“值得过”。这个描述可能会被那些认为不会有不值得过的生活的人们所忽视。但是，像许多别的人们一样，我相信会有这样的一些生活。最后，我展开了“值得过”这一短语的日常用法。如果两个人中的一个具有较低的一种生活质量的话，我在这个意义上称其生活“较不值得过”。

当我们考虑到未来的人的时候，有两个问题必须回答：

（1）如果我们使某个人存在，其将有一个值得过的生活，我们从而有益于这个人吗？

（2）如果我们的某个行动是其存在原因中一个遥远但必须的部分，我们还有益于这个人吗？

这是些困难的问题。如果我们对两者的回答都是肯定的，我将说我们认为使某人存在能够带来益处。

一些人对（1）的回答是肯定的，但是对（2）的回答是否定的。这些人之所以给出了第二种回答，是因为他们在通常的意义上使用"带来益处"这个词。正如我在第25节中所论述的，出于道德上的一些目的我们应该展开"带来益处"的用法。如果我们对（1）的回答是肯定的，那么我们对（2）的回答就应当是肯定的。

许多人对这两个问题的回答都是否定的。这些人可能会说："倘若我们没有做我们所做过的事的话，对这个人而言本会更糟，如果这一点为真，那么我们就有益于某个人。如果我们没有使某个人存在的话，那么对这个人而言本来就不会更糟。"

我认为，尽管对这两个问题都给出否定的回答是可以得到辩护的，但是对这两个问题都给出肯定的回答也是可以得到辩护的。为那些怀疑这第二个信念的人们，我在本书写下了"附录"G。既然我认为，主张和否定使某人存在能够带来益处都是可以得到辩护的，我将讨论这两个观点的蕴含。

考虑一下

那个14岁大的女孩。这个女孩选择生一个孩子。因为她还太小，她提供给她的孩子一个糟糕的人生起点。尽管在这个孩子的一生中这将产生糟糕的影响，但可以预测的是他的生活是值得过的。如果这个女孩再等待数年的话，她本会有一个不同的孩子，她本可以为他提供一个更好的人生起点。

鉴于这样的事例变得很普遍，它们引发出一个实践性的问题。[7]它们也引发出一个理论性的问题。

假设我们劝说过这个女孩应该等待。我们主张说："如果你现在生孩子，你将很快为此感到后悔。如果你等一等，对你来说会更好。"她回答说："这是我自己的事。即使我是在做一件对我来说将会更糟的事情，我有权做我想做的事。"

我们回应说："这并非完全是你的事。你不仅应当想到你自己，还应当想到你的孩子。如果你现在生他的话，对他而言将更糟。如果你以后生他，你将给予他一个更好的生活起点。"

我们没有能够说服这个女孩。她14岁的时候生养了一个孩子，而且就像我们预测到的那样，她给予他的是一个糟糕的人生起点。我们主张她的决定对她的孩子而言是更糟的这一点正确吗？如果她等待了的话，这个特定的孩子就永远不会存在了。况且，尽管他的生活起点糟糕，但是他的生活是值得过的。首先假设我们并不认为使某人存在能够带来益处。我们应当问："如果某个人过着一种值得过的生活，对这个人而言，这比倘若这个人本来永远不会存在更加糟糕吗？"我们的回答一定是否定的。接着假设我们认为使某人存在能够带来益处。照此观点，这个女孩的决定有益于她的孩子。

照这两种观点，这个女孩的决定对她的孩子而言都不是更糟的。当我们看清这点的时候，我们改变关于这个决定的看法吗？我们不再认为，如果这个女孩等一等的话，她能够给予她的孩子一个更好的人生起点本会更好吗？就像那些考虑这个事例的绝大多数人那样，我继续持有这一观点。但是我们不能够用我所提出的那种自然方式辩护这个信念。我们不能够主张说，这个女孩的决定对其孩子而言是更糟的。对她的决定的诘难是什么？这个问题之所以产生，是因为在不同的后果中出生的人会不同。因而我将称此为非同一性问题。[8]

可能有人说：

在某种意义上，这个女孩的决定对她的孩子而言是更糟的。在努力劝说这个女孩现在不要生孩子的过程中，我们可以使用"她的孩子"这个短语和代词"他"涵盖她可能生养的任何一个孩子。这些词语无需指称一个特定的孩子。我们能够正当地主张："如果这个女孩现在并

不生她的孩子，而是等到以后再生养的话，他将不是那同一个特定的孩子。如果她后来生下他，他将是一个不同的孩子。”通过以这种方式使用这样一些语汇，我们能够解释为什么这个女孩如果再等一等生孩子会更好。我们可以主张说：

（A）对这个女孩所作决定的诘难是：对她的孩子而言，它很可能是更糟的。如果她等待的话，她很可能会为他提供一个更好的人生起点。

尽管我们可以正当地提出这一主张，但是我们并不解释对这个女孩的决定的诘难。这一点在她生养下她的孩子之后，变得十分明显。短语“她的孩子”现在自然是指称这个特定的孩子。而且这个女孩的决定对这个孩子而言并不更糟。尽管在某个意义上（A）为真，但是（A）并不诉诸一个人们所熟悉的道德原则。

按照我们所熟悉的那些原则中的一个原则，如果某个人的选择对另外任何一个特定的人而言更糟，或者有悖于另外任何一个特定的人的利益，就是对这个人的选择的诘难。如果我们主张说这个女孩的决定对她的孩子而言更糟，那么，我们就不可能是在主张它对某个特定的人而言更糟。就女孩的孩子来说，我们不能主张，她的决定对他而言更糟。我们必须承认，在主张（A）中，语汇“她的孩子”并不指称她的孩子。（A）并不指称那些生活着的特定的人们中任何一个人的好或者糟。（A）诉诸一个新原则，必须予以解释和提供合理性证明。

如果说（A）看起来诉诸一个人们所熟悉的原则，这是因为它有两个意思。再举一个例子。如果在许多战斗中，一位将军总能够使自己属于取胜的一方，他就表现出有军事才能。但是做到这一点有两种途径。他可能赢得胜利。或者他可能在即将失败的时候扭转局面。如果一个将军只是在第二种意义上总是能够使自己属于取胜的一方，他就没有表现出任何军事才能。

（A）所诉诸的是什么原则？我们应当以这样一种方式来陈述该原则，这种方式能够表明该原则所诉诸的那类选择。这些是相同数目选择，这影响到未来的人的同一性，但是并不影响到他们的数目。我们可以提出

相同数目质量主张，或曰Q主张：如果在两种可能的结果中相同数目的人都会生活着，如果那些生活着的人变得更糟，或者所具有的生活质量比那些本来会生活在其中的人们的生活质量低的话，就会是更糟的。

这个主张是令人信服的。而且它蕴含着有关那个14岁的女孩我们所持的信念。她现在生的那个孩子很可能比她本会以后生的一个孩子变得更糟，因为这另一个孩子本会有一个更好的生活起点。如果此点为真，Q蕴含着这是两个可能的结果中更糟的那个。Q蕴含着如果这个女孩等到以后再生一个孩子本会更好。

我们可能从就这个女孩事实上的孩子来说如果他永远不存在本来会更好这个主张退缩回来。但是，如果我们早先主张过如果这个女孩等一等本会更好，那么这就是我们所必须主张的。我们不能前后一致地既肯定、随后又否定同一个主张。如果（1）在1990年倘若这个女孩等一等，到以后再生孩子本将会更好的话，那么（2）在2020年倘若她已经等了并且到以后生养了孩子的话，本已经会更好。而且（2）蕴含着（3）：如果那个存在了的孩子并不是她事实上的孩子的话，本来会更好。如果我们不能接受（3），那么我们必须拒斥（1）。

我提出，经过反思，我们能够接受（3）。我认为，如果我是这个女孩的那个事实上的孩子，我会接受（3）。（3）并不蕴含着我的存在是糟的，或者在道德上是内在地不可取的。该主张只是，既然一个后来出生的孩子本来很可能会具有一种比我的生活更好的生活，那么，如果我的母亲等待了，而且后来生了一个孩子的话，本会更好。这个主张无需蕴含我应该合乎理性地为我的母亲生下我而感到后悔，或者她应该合乎理性地为此感到后悔。既然，如果她等待了本会更好的话，她大概应该具有某种道德上的后悔。而且，她使结果对她本身也更糟很可能也是真的。但是，如果予以全面的考虑，即便这是真的，它并不表明她应该合乎理性地后悔她的行动。如果她爱我——她事实上的孩子，这就足以堵死有关她没有这样的悔意就是非理性的这

一主张。[9]即使当它蕴含像（3）一样的主张的时候，我断言我们也能够接受Q。

尽管Q令人信服，但是它并未解决非同一性问题。Q只涵盖那些在不同的结果中同样数目的人会生活着的事例。我们需要一个主张，涵盖那些在不同的结果中不同数目的人会生活着的事例。非同一性问题能够出现于这些事例中。

因为Q受到限制，可以用不同的方式证明其正当性。有数个原则蕴含Q，但是当应用于不同数目选择的时候，却有冲突。我们需要决定，这些原则中的哪一个，或者哪一组，我们应当予以采纳。我们应该予以采纳的被称为X理论。X理论将解决不同数目选择中的非同一性问题。而且X理论告诉我们Q应当如何得到正当性证明，或者如何得到更充分的解释。

在那个14岁女孩的事例中，我们并未被迫诉诸Q。还有其他一些我们本可诉诸的事实，诸如对其他人的影响等。但是该问题能够以一种更纯的形式出现。

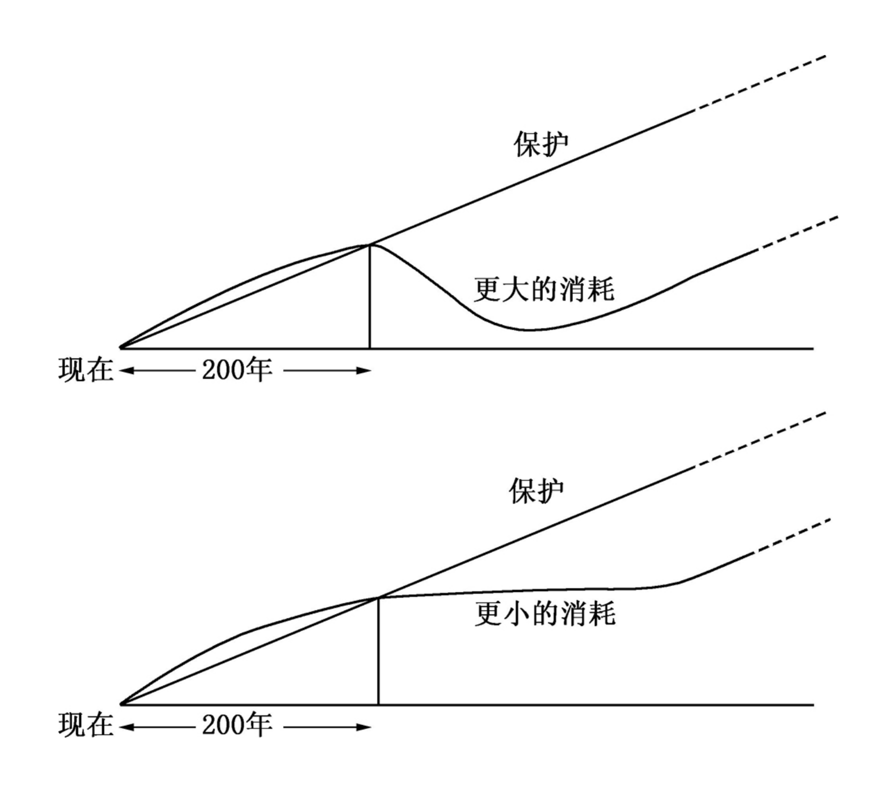
123.降低生活质量何以不会对任何人更糟

假设我正在两个社会或者经济政策之间作选择。而且假设，照两个政策中的一个政策，生活标准会在21世纪中略有提高。这个影响蕴含着另一个影响。无论我选择哪一个政策相同的一些特定的人们将在更远的未来存在这一点并不为真。有鉴于这样的两个政策对我们具体生活的影响，按不同的政策人们会与不同的人结婚这一点会历时地越来越真。而且，即使在同样的那些婚姻中，孩子们越来越历时地在不同的时刻成胎。正如我已经论述过的，那些孩子成胎早1个月或者晚1个月，事实上就会是些不同的孩子。既然我们的两个政策之间的选择会影响到以后受孕的时间性，那些以后出生的人中的一些人的存在就归功于我们对两个政策中的一个的选择。如果我们选择的是另一个政策的话，这些特定的人们就永远不会存在了。而且，那些以后出生

的、其存在归功于我们所作的选择人们的比例，会像池塘中的涟漪一样不断地增大。我们能够令人信服地假定，经过一两个世纪之后，在我们的共同体中那些无论我们选择哪一个政策他们都会存在的人们不会再有人活着。（思考一下这个问题可能有所帮助：我们有多少人可以正当地主张，“即便火车和汽车从未发明出来，我仍然会出生？”）

这何以产生出一个问题？考虑一下

消耗。作为一个共同体，我们必须选择是消耗还是保护某些种类的资源。如果我们选择消耗，生活质量在下两个世纪中比倘若我们选择了保护的话会有略微的提高。但是生活质量在后来的许多世纪中会比倘若我们选择了保护的话要低得多。这会是因为在这个时期之初，人们会不得不为我们所消耗掉了的资源寻找替代品。对这个事例的两个版本进行区分是值得的。不同政策的后果会像下图所示。



我们决不会如此具体地知道这些就是两个政策的后果。但这绝不是对这个事例的诘难。相似的后果有时候会是可以预测的。这个想象出的事例人为地简单也并不重要，因为这只是澄清相关问题。

假设我们选择消耗，而且这具有如图所示的两种后果中的任何一个。我们的选择对任何人而言都更糟吗？

因为我们选择了消耗，数以百万计的人们在数个世纪中的生活质量要低得多。这种生活质量不是比现在的生活质量低得多，而是比倘若我们选择了保护的情况下的生活质量低得多。这些人的生活是值得过的；而且，如果我们选择了保护的话，这些特定的人根本就不会存在。假设我们并不假定使某人存在能够带来益处。我们应当问，“如果特定的一些人过着值得过的生活，对这些人而言这比他们本来决不会存在更糟吗？”我们的答案必定是否定的。接着假设我们的确假定使某人存在能够带来益处。既然这些未来人的生活将是值得过的，而且如果我们选择了保护，他们根本就不会存在的话，那么，对这些人而言，我们选择消耗非但不是更糟，反而是有益于这些人。

照这两种回答，我们的选择对这些未来的人将并不更糟。况且，当我们理解该事例的时候，我们知道这的确如此。我们知道，即使它极大地降低了数个世纪的生活质量，我们的选择对任何一个活着的人都将不是更糟的。

这造成任何道德上的差异吗？有三种观点。它或许造成所有的差异，或者造成某种差异，抑或不造成什么差异。对我们的选择，或许没有什么诘难，或者有某种诘难，抑或有同样强烈的诘难。

一些人认为，糟糕的东西必定是对某人而言是糟糕的。照此观点，对我们的选择没有什么诘难。既然它对没有什么人而言是糟糕的，我们的选择不可能有糟糕的后果。生活质量的大幅度降低没有提供什么不选择消耗的道德理由。

某些学者接受这个结论。[10]但这个结论却是非常不令人信服的。在我们考虑此类事例之前，我们可能采纳有关糟糕的东西必定是对某人而言是糟糕的这一观点。但我认为，消耗事例所表明的是，我们必须拒斥这个观点。生活质量的大幅度降低必定提供不选择消耗的某种道德理据。这为那些考虑此类事例的绝大多数人所接受。

如果这是我们所认为的，我们应当问两个问题：

（1）不选择消耗的道德理由是什么？

（2）生活质量的这种降低将不会对什么人而言更糟的这一点造成一种道德差异吗？如果它对特定的一些人而言是更糟的话，这个影响会是更糟，具有更大的道德权重吗？

对于我们回答问题（1）和其他相似问题的需要，我称为非同一性问题。这个问题之所以产生，是因为更遥远的未来的人们的同一性可能非常容易受到影响。一些人认为这个问题纯粹是一个遁词（quibble）。这种反应得不到正当性证明。该问题的产生是因为与我们的生殖系统相关的一些肤浅事实。但是，尽管它是以一种肤浅的方式产生的，却是一个真实的问题。当我们在两个社会政策或者经济政策之间作出选择的时候，就像我所表述的那一类，无论我们选择什么，在遥远的未来同一些人将存在这一点并非为真。因而，像消耗这样的一种选择将有悖于未来的人的利益并非为真。我们并不能够假借这为真来取消这个问题。

如果我们诉诸Q，我们部分地回答问题（1）。照此主张，如果数目是同样的，倘若那些生活着的人具有比那些本会生活着的人更低的生活质量，就会是更糟的。但是该问题能够出现在那些不同的结果中会有不同数目的人的事例中。为了涵盖这些事例，我们需要X理论。只有X理论将解释Q如何应当得到正当性的证明，以及如何提供解决我们这个问题的一个完全方案。

124.为什么诉诸权利不能解决该问题

我们能够通过诉诸人们的权利来解决我们的问题吗？再考虑一下那个14岁的女孩。她生孩子如此早，以至于赋予孩子一个糟糕的生活起点。或许有人主张："对这个女孩所作决定的诘难是，她侵犯了她孩子享受良好生活起点的权利。"

即使这个孩子有这个权利，也根本不会实现。等到这个女孩是成熟女性的时候，她根本不会有这个孩子。一些人会主张说，既然这个孩子的权利根本不会实现，不能说这个女孩侵犯了他的权利。诘难者或许回应说：“如果我们知道某个人将具有一个不能实现的权利，还使这个人存在是不正确的。”这能够是对这个女孩所作决定的诘难吗？[11]

一些年前，有一位英国政治家对前一年青少年怀孕减少的事实表示欢迎。一个中年男子满怀愤怒地给《泰晤士报》写了一封信。他就是在妈妈只有14岁的时候出生的。他承认，由于他妈妈太年轻，他的早年生活对他和他母亲两者来说都相当艰难。但是他现在的生活是非常值得过的。那位政治家是在暗示，如果他根本没有出生的话本会更好吗？这种暗示在这个中年人看来令人难以容忍。

那位政治家提出了这一点，有这么一种蕴含。照这位政治家的观点，这个人的母亲如果再等几年生孩子的话本来会更好。我认为，我们应当接受这个观点。但是，我们能够凭借这样一种主张——这个愤怒的人享有一种没有得到实现的权利——而具有可信度地解释这个观点吗？

我认为我们不能。假设我拥有隐私权。我请求你与我结婚。如果你接受了我的请求，你的行动并不因侵犯了我的隐私权而成为错误的。既然我对你能这样做而感到欣喜，那么，具体到对你而言，我则放弃这个权利。同样的主张也适用于那位愤怒地给《泰晤士报》写信的人。照上述提法，这个人享有生自一位成熟女性的权利，这位成熟的女性会为他提供一个良好的生活起点。这个人的母亲的行动之所以是错误的，是因为她使他存在，却带有一种不能实现的权利。但是这个人的书信表明，他生活得很好。他否认他的母亲之所以行动错误是因为她对他的所作所为。如果我们主张说，因为他享有一种不能实现的权利，她的行动错误的话，他就会说，“我放弃这个权利”。这会从根本上瓦解我们对他母亲的诘难。

如果这个人的母亲等待了的话，本来会更好。但这不是因为她对她的事实上的孩子所做的，而是因为她本来会为成熟的时候再生出的任何一个孩子所做的。诘难必定是这样：如果她等待了的话，她本来会为某个别的孩子提供一个更好的生活起点。

现在回到那个消耗的事例。假设我们选择了更大的消耗。两个多世纪之后，生活质量要比我们选择了保护的话要低得多。但是那个时候生活的人将拥有几乎与我们在21世纪中的平均水准相当的生活质量。这些人享有诘难者能够诉诸的权利吗？

可能有这样的主张，即这些人有权享有我们所消耗掉的那些资源中的份额。但是人们无权享有某种特定资源的份额。假设我们消耗了某种资源，我们的后继者们缺乏这种资源，但是我们发明了将能够使我们的后继者们拥有与我们相同的一系列机会的技术。那么对我们所做的就不会有什么诘难。最可能为人们所主张的诘难是，每一代人都有权享有平等的机会或者同等生活质量。[12]

如果我们选择更大的消耗，那些生活在两个多世纪之后的人们将有的机会比生活在早些或者晚些世纪中的人们的机会要少，生活质量要低。如果人们有权享有平等机会和同等生活水准，诉诸这些权利可能提供某种对我们的选择的诘难。那些生活在两个多世纪之后的人们根本不会享有更大的机会或者更高的生活质量。如果我们作了别的选择的话，这些人们可能根本就不会存在。既然他们的权利不会得到实现，我们就不可能侵犯到他们的权利。但是，如前所述，会有人诘难说，我们使那些具有不能实现的权利的人们存在。

这是一个很好的诘难，这点并不清楚。如果这些人知晓那些事实的话，他们不会痛惜我们所做的那些行动。如果他们高兴地活着的话，他们的反应可能就会像那位给《泰晤士报》写信的人一样。他们可能会放弃他们的权利。但是，既然我们不能够假定这就是他们所有人的反应，那么，诉诸他们的权利可能提供了某种对我们的选择的诘难。

这样的诉诸能够提供一种对我们选择更小的消耗的诘难吗？在这种情况下，那些在两个多世纪之后生活的人们将享有比我们现在高得多的生活质量。我们能够主张这些人享有甚至更高生活质量的权利吗？我相信，按照任何一个令人信服的权利理论，答案都会是否定的。

它将有助于设想去掉那个非同一性问题。假设我们的生殖系统非常地不同。假设，无论我们遵循哪一个政策，这些相同的人都会在两个多世纪之后生活。对我们的选择的诘难会是这样的：因我们自己和我们子女的小利益之故，我们阻碍了未来的人获得大得多的利益。既然这些未来的人的境况要比我们更好，那么我们就不是在行事不公。对我们的选择的诘难就会不得不诉诸功利原则。

这个诘难会诉诸权利吗？仅当，像戈德温那样，把功利主义表现成一种有关权利的理论时才如此。照戈德温的观点，每一个人都有权享有功利原则中应有之义所蕴含的我们应当得到的东西。那些相信权利的绝大多数人会拒斥这个观点。许多人把权利解释为约束或者限制功利原则的东西。这些人主张，侵犯某些权利是错误的，即便这极大地增加了抵消掉负担之后的总利益净量。照这样的一种理论，功利原则获得某种权重。既然这样的一种理论不是功利主义的，该原则最好称作善行原则。这个原则是这样的一种理论的一个组成部分，而我们享有某些权利的主张则是这个理论的另一个不同的组成部分。我将假定，如果我们相信权利，这就是我们所接受的那种道德理论。

回到我们设想去掉的非同一性问题的那个事例当中。如果我们拒斥戈德温的观点，我们凭诉诸那些将在更遥远的未来生活的人们的权利不能诘难我们选择更小的消耗。我们的诘难会诉诸善行原则。诘难会是这样的：因自己的和子女的小利益之故，我们否定了比我们境况更好的那些人们大得多的利益。在称之为诘难方面，我无需主张它表明我们的选择是错误的。我只是在主张，既然我们否定了这些人的大得多的利益，这提供了某种不要作出这个选择的理由。

如果我们现在恢复我们实际上的生殖系统，这个理由便消失了。考虑一下那些将在两个多世纪之后生活的人们。我们选择更小的消耗并未否定这些人的任何利益。如果我们选择了保护的话，也根本不会有益于这些人们，因为他们本就决不会存在。

当我们假设非同一性问题的时候，我们不作这样选择的理由不是由诉诸人们的权利而是由诉诸善行原则得到解释的。当我们重拾非同一性问题的时候，这个理由便消失了。既然这个理由所诉诸的是善行原则，该问题所表明的是这个原则不充分，必须予以修正。我们需要一种对善行的更好说明，或者我所称的X理论。

我们道德理论的一个部分诉诸善行；另一个部分诉诸人们的权利。因而我们不应当期待诉诸权利能够填补我们不充分的善行原则中的空白。我们应当期待的是，正如我所已经主张的，诉诸权利不能够完全解决非同一性问题。[13]

125.非同一性这个事实造成一种道德差异吗?

在努力修正我们的善行原则方面——在努力发现X理论方面，我们必须考虑到在不同的结果中会存在不同数目的人的那些事例。在我们转向这些事例之前，我们能问这样一个问题，就是关于我所提及的另一个问题我们相信什么。我们对消耗的选择不会对什么人来说是更糟的。这会造成一种道德上的差异吗?

我们也许能够记得，曾几何时我们关心那些对未来的世世代代的影响，但却忽略了非同一性问题。我们可能想过，一个像消耗这样的政策会有悖于未来的人的利益。当我们看清此为假的时候，我们变得较少关心对未来的世世代代的那些后果了吗?

当我看清该问题的时候，我并不变得较少关心起来。对其他许多人而言情况亦然。我将说，我们接受无差异观点。

值得考虑一下另一个不同的例子：

医疗方案。有J和K两种罕见的病症，不经过特别的检验无法发现。如果一位孕妇罹患J病症，将导致她所怀的胎儿出现某种残疾。简单的治疗就会避免这种不利影响。如果一位女性在孕育一个胎儿的时候患上了K病症，将导致这个孩子出现那个相同的特定残疾。K病症是不可医治的，但是总在2个月内消失。接着假设我们计划了两个医疗方案，但是只有资助一个方案的基金，所以其中的一个要被取消。在第一个方案中，数以百万计的妇女在孕期会得到检测。那些被发现患上了J病症的人会得到治疗。在第二个方案中，数以百万计的妇女在力图怀孕的时候得到检验。那些被发现患有K病症的人将会得到至少延期2个月怀孕的警告，其时这个不可医治的病症将会消失。最后假设我们能够预测这两个方案会在同样多的事例中取得结果。如果有孕期检测，每年将有1000个孩子出生正常，而不是身有残疾。如果有孕前检测，每年将有1000个孩子出生正常，而不是出生1000个与之不同的残疾孩子。

这两个方案同等值得推行吗？让我们仔细注意差异之所在。作为两个方案中任何一个方案的结果，每年1000对夫妻都会各自生下一个正常的而不是身有残疾的孩子。照两个方案，这些会是些不同的夫妻。但是，既然数目会是相同的，对那些父母和其他人的影响在道德上会是等效的。如果有一种道德差异的话，这只能够是在对孩子们的影响方面。

接着要注意到，在判断这些影响方面，我们需要对胚胎的道德地位不持任何观点。我们可以假设，在任何一种检测开始之前要过上1年的时间。当我们在这两个方案之间进行选择的时候，那些孩子们中还没有任何一个成人胎。而所有那些成人胎的都将会成为成年人。因而，我们正在考虑的后果不是及于现在的胚胎，而是及于未来的人。接着假定，所谈及的残疾尽管并非微不足道，却也没有严重到引发生活是否值得过的怀疑。即使他们生下来可能有悖于我们的利益，但是对那些生来带有这种残疾的人而言却不然。

既然我们无力支撑两个医疗方案，我们应当撤销哪一个？按照一种描述，两者会有相同的后果。假设J病症和K病症是这种残疾的惟一原因。现在的发生率是每年出生的那些人中的2000人。两种方案中任何一个方案都会使发生率减半；比率就会跌至每年1000人。差异就在此。如果我们决定取消孕期检测，对那些后来生有残疾的人来说，如果不是因为我们的决定的话，他们本会得到治愈是真的。我们的决定

对所有这些人而言都是更糟的。如果相反我们决定取消孕前检测，以后将有同样多数目的人生有这种残疾。但是对这些人而言，如果不是因为我们的决定的话，他们本会得到治愈却不会为真。这些人的存在归因于我们的决定。如果我们没有决定取消孕前检测的话，这些残疾孩子的父母本不会生下他们。他们以后本会有一些与之不同的孩子。既然这些残疾孩子们的生活是值得过的，那么我们的决定对他们中的任何一个而言都将不是更糟的。

这造成一种道德差异？还是这两个方案同等值得推行？有多少未来的生活由那些正常的而不是有残疾的孩子去过在道德上是重要的？还是这些生活是否由与之相同的一些人过活也重要？

我们应当为这个事例增加一个细节。如果我们决定取消孕期检测，那些以后生有残疾的人们可能知道，如果我们作出的是不同的决定的话，他们本来是可以治愈的。这样的认识可能会使他们的残疾更加难以忍受。从而，我们应当假定，尽管它不是被精心考虑地予以取消，但是这些人不会知道这个事实。

带着这个添加的细节，我判定这两个方案是同等值得推行的。我知道有一些人不接受这个主张；但是我知道更多的人接受这个主张。

我的反应不仅仅是一种直觉。它是我按照以下推理得到的结论。无论哪一个方案被取消，以后将会有同样多数目的人身患这种残疾。这些人在有赖于我们的选择的那两种结果中会是不相同的。而且有一个主张只适用于这两组残疾人中的一组。但是他们不知道有一组人本来是可以得到治愈的这个事实。我从而发问：“如果将有一些身患某种残疾的人，他们是残疾的这个事实是糟糕的。如果他们的残疾本来是可以得到治愈的话（他们并不知道这一点），这岂不会更糟吗？”如果这个事实使这些人比那些残疾本不会得到治愈的人们的境况更糟的话，这会是更糟的。但是这个事实并不具有这样的后果。如果我们决定取消孕期检测，将会有一组残疾人。如果我们决定取消孕前检测，将会有一组与之不同的残疾人。第一组中的人不会比第二组中的人境

况更糟。既然如此，我判定这两个结果在道德上是等效的。鉴于该事例的细节，一组人而不是另一组人本来会得到治愈在我看来是不相关的。

如果治愈这一组人就会减少这个残疾的影响范围的话，这个事实本会是相关的。但是，既然我们只有一个方案的资金，这就不为真。如果我们选择要治愈第一组人，后来将有同样多数目的人生有这种残疾。既然治愈第一组人不会减少将残疾的人的数目，只有当这组人具有更强烈的治愈主张的时候，我们才应该选择治愈这一组人。而他们并没有一种更强烈的主张权。如果我们本会治愈第二组，他们本会具有同等的治愈主张权。如果我们选择去治愈第一组，那么他们只不过是比第二组更幸运而已。既然他们只是更幸运而并没有更强的主张治愈的权利，所以我并不认为我们应该选择治愈他们。既然即便我们选择治愈这些人，也将不会减少将要残疾的人的数目这一点也为真，那么我断言这两个方案同等值得推行。如果孕前检测会在更多一些的事例中取得结果的话，我就会判定它是更好的方案了。[14]

这与我对我们的消耗选择的反应相匹配。我认为，如果以后会有一种低得多的生活质量的话，就会是糟的。而且我认为，如果我们选择了保护，如果那些以后生活的人本身也会存在的话，就不会是更糟的。如果糟的影响在这个意义上对任何特定的人们而言是更糟的话，那么它就不会是更糟的。在考虑这两个事例方面，我都接受无差异观点。别的许多人也是如此。

我表述了其中我和其他许多人都接受无差异观点的两个事例。如果我们接受这个观点是确当的，这可能具有重要的理论性蕴含。这有赖于我们是否相信，如果我们造成某个将有值得过的生活的人的存在，我们从而有益于这个人。如果我们相信这一点，我们仍然不能陈述无差异观点的蕴含，因为这些将依赖于我尚未讨论的决定。但是假

设我们相信使某人存在不能有益于这个人。如果这就是我们所相信的，而且我们接受无差异观点的话，那么所蕴含的如下：

我已经提出，我们应当诉诸

Q：如果在两种可能的结果中相同数目的人都会生活着，如果那些生活着的人变得更糟，或者所具有的生活质量比那些本来会生活在其中的人们的生活质量低的话，就会是更糟的。

接着考虑一下：

对人产生影响的观点或曰V观点：如果人们受到变得更糟的影响的话，这就将是更糟的。

在同一些人选择中，Q与V相合。当我们在考虑这些选择的时候，那些生活着的人在两个结果中都是同一些人。如果这些人处境更糟，或者具有一种更低的生活质量的话，他们则受到变得更糟的影响，反之亦然。[15]既然Q和V在此处彼此相合，我们诉诸哪一个将不造成什么差异。

这两个主张只有在相同数目的选择中才呈现冲突。假设我们接受无差异观点。在考虑这些选择方面，我们那么将诉诸Q而不是V。如果我们选择消耗，这将降低更远的未来中生活质量的水准。根据Q，我们的选择具有糟糕的结果。但是，由于有关同一性的那些事实，我们的选择将不会对什么人来说是糟糕的。V并不蕴含着我们的选择有一种糟糕的后果。如果这个后果对特定的人们而言是更糟糕的，那么这个后果是更糟糕吗？如果我们所诉诸的是V而不是Q，我们的回答会是肯定的。但是，既然我们相信无差异观点，我们作出否定的回答。我们认为V在此给出了正确的回答。而且V在医疗方案中给出了正确的回答。Q描述了那些我们认为是糟糕的后果。而且我们认为，这些后果根据V是否也是糟糕的，并不造成什么道德差异。V在按照我们的观点而言不应当作区分的地方作了道德区分。

在同一些人选择中，Q和V彼此相合。在相同数目选择中，此处这些主张相冲突，我们采纳Q而不是V。当我们作这两类选择的时候，我们因而将不运用V。

还剩下有不同数目选择，此点Q并不涵盖。我们此处将需要X理论。我尚未讨论X应当主张什么。但是我们能够预测如下。在相同数目选择中X将蕴含Q。

我们还可以预测X将具有与V相同的关系。在同一些人选择中，X与V将彼此相合。我们诉诸哪一个此处并不造成任何差异。这些是我们的道德思维最为关注的选择。这解释了V的令人信服之处。道德的这一部分，那有关善行或人的福祉的部分，通常用我所称的对人产生影响的术语加以思考。我们诉诸人们的利益——诉诸对我们的行动所影响到的那些人来说或好或糟的东西。即使在我们发现了X理论之后，我们或许在绝大多数情况下继续诉诸V，只是因为它为人们更加熟悉。但是在某些情况下，X和V将出现冲突。当我们在作出相同数目选择和不同数目选择的时候，他们可能冲突。而且，每当X和V冲突的时候，我们将诉诸X而不是V。我们将相信，如果某个后果根据X是糟糕的，根据V它是否也是糟糕的，并不造成什么道德差异。就像以前一样，V在照我们的观点来说不应当作任何区分的地方作了区分。V就像有关奴役白人或者剥夺成年男性的选举权是错误的这一主张一样。我们于是将断言，道德的这个部分，那有关善行和人的福祉的部分，不能够以对人产生影响的术语来解释。它的那些基本原则与我们的行动对所影响到的那些人而言是好还是糟并不有关。X理论将蕴含着，如果一个后果对人们而言是糟糕的，它就是糟糕的。但这将不会是为什么这个后果是糟糕的这一点的理由。

接着要谨记，这些主张假定使某人存在不能带来益处。这个假设是可以得到辩护的。如果我们作这样的假定，这些主张表明许多道德理论需要修正，因为这些理论蕴含着，无论我们的行动对所影响到的那些人来说是好还是糟，都必定造成一种道德上的差异。[16]而且我们

可能需要对某些常见情况所抱的信念加以修正。堕胎或许就是一个例子。但是我们道德思维的大部分会保持不变。许多重大的关系只保持在特定的人们之间。其中包括我们与我们作出承诺的那些人之间的关系，或者与我们心怀感激的那些人之间的关系，抑或与我们的父母、学生、病人、客户，以及（如果我们是政治家的话）我们所代表的那些人们之间的关系。我的评述并不适用于这样的一些关系，或者它们所产生的特殊义务。我的评述只适用于我们的善行原则：即只适用于要有益于他人和防止他们受到伤害的一般道德理由。

既然我的评述只适用于这个原则，而且我们将只会在某些情况下改变我们的观点，这个观点改变可能看来并不重要。这并非如此。再次考虑一下这样一个（过于堂皇）的类比：在通常情况下，我们能够接受牛顿定律，但并非在所有情况下。而且我们现在是接受一个不同的理论。

126.造成更遥远的未来中的一些可以预见的灾难

在本节中，我不是沿着我的论证主线继续往下求索，而是讨论一个小问题。在像消耗这样的事例中，我们凭借诉诸人们的权利不能够完全解决非同一性问题。在这个事例的一种变体中——其中我们的选择造成一场灾难——也是如此吗？既然这是一个小问题，除了那些并不相信消耗具有糟糕的影响的人们之外，本节完全可以略过。考虑一下

冒险政策。作为一个共同体，我们必须在两个能源政策之间作出选择。在至少3个世纪之内两者都完全是安全的，但是其中的一个在更遥远的未来会有某些潜在的危险。这个政策涉及到在这样一些区域填埋核废料，在接下来的为数不多的世纪中这些区域没有发生地震的危险。但是，鉴于这种废料在长达数以千计的年份里都具有放射性，在遥远的未来将有一些危险。如果我们选择这个冒险政策，下个世纪中的生活水准将会高一些。我们的确选择了这个政策。结果，许多世纪之后出现了一场灾难。由于地球表面发生地质变化，一场地震造成核废料的核辐射泄漏，数以千计的人因此丧生。这些人本将享有值得过的生活，但是这场灾难造成他们罹难。我们可以假定，这场核辐射只影响到那些在核辐射泄漏之后出生的人们，

而且令他们身患不治之症，将会在他们大约40岁的时候夺去他们的生命。这种疾病在夺去人的生命之前不会产生什么后果。

我们在这两个政策之间的选择将影响到以后人们所享受到的生活的那些细节。按我们上述的解释方式，我们的选择从而将影响到以后到底谁将生活着。许多世纪之后，无论我们选择哪一个政策都会出生的那些人，在我们的共同体中不会还有什么人生活着。由于我们选择了冒险政策，数以千计的人后来丧生。但是，如果我们本来选择的是可供选择的安全政策的话，这些特定的人们就根本不会存在，取代他们位置的将是一些与之不同的人。我们选择冒险政策对任何什么人来说是更糟吗?

我们应当问，"如果人们过着值得过的生活，即使他们在某场灾难中丧生，这比他们根本就不曾存在更糟吗？"我们的回答必定是否定的。尽管它造成了一场可以预见的灾难，但是我们选择冒险政策将不会对什么人来说是更糟的。

一些人可能说，我们选择消耗并不具有糟糕的后果。但关于我们对冒险政策的选择却不能够如此主张。既然这个选择造成一场灾难，它显然具有糟糕的后果。但是，我们的选择将不会对以后生活的人们中的任何一个人来说是糟糕的或者更糟的。这个事例迫使我们拒斥这样一个观点，即如果一个选择不会对什么人而言是糟糕的话，这个选择就没有糟糕的影响。

在这个事例中，非同一性问题可能看起来更容易解决。尽管我们的选择对那些遭受灾难的人们来说不是更糟的，但是可能可以说，我们伤害到了这些人。而且，诉诸人们的权利在此可能取得成功。

我们该当因为伤害到他人而受到责备，即便这对他们而言并非是更糟的。假设我驾车鲁莽，结果造成车祸，令你失去一条腿。1年之后，战争爆发。如果不是因为你失去一条腿的话，你本会应征入伍，在战斗中阵亡。我驾车鲁莽反而救了你一命。但我仍然该受道德的谴责。

这个事例提醒我们，在确定什么予以谴责的时候，我们一定不但要考虑到那些事实上的影响，还要考虑到那些可以预见到的影响。我知道我的鲁莽驾驶或许伤害到别人，但是我不会知道事实上这会救你一命。这一区分或许可以适用于我们对冒险政策的选择。假设我们知道，如果我们选择这个政策，就可能在遥远的未来造成许多人意外死亡。但是我们已经把非同一性问题抛到了脑后。我们错误地认为，无论我们选择哪一个政策，同一些人将在以后生活。我们从而认为，我们选择冒险政策可能大大违背某些未来人的利益。如果我们相信此点，我们的选择就会受到批评。我们会该当为做了我们信以为可能大大违背他人利益的事而受到谴责。即使我们的信念是虚假的，这个批评仍然能够站住脚——这就像，即使我的鲁莽驾驶将在事实上救你一命，我同样要受谴责。

假设我们不能找到X理论，或者X理论看起来比对做可能大大违背他人利益之事的诘难更不可信，那么如果我们向那些将要决定是否增加使用核能源的人们隐瞒非同一性问题可能更好一些。如果这些人虚假地认为，这样一个造成灾难的政策可能大大违背那些将在遥远的未来生活着的人们中的一些人的利益，这也许更好一些。如果这些人具有这个虚假的信念，他们也许更有可能得出一些正确的结论。

我们已经失去了这个虚假的信念。我们意识到，如果我们选择冒险政策，我们的选择对那些在那场灾难中丧生的人们而言将不是更糟的。注意，这不是一种幸运猜测。这不像预测说，如果我造成你失去一条腿，在以后这将使你免遭在战壕中丧生的命运。我们知道，如果我们选择冒险政策，这可能在遥远的未来造成许多人丧生。但是我们也知道，如果我们本来选择了安全政策的话，那些丧生的人就根本不会存在。既然这些人的生活将是值得过的，我们知道我们的选择对他们而言并非是更糟的。

如果我们知道这点，我们就不能够被比作一个鲁莽的司机。对我们所作选择的诘难是什么？当我们知道我们的行动对受到伤害的那些

人而言将并非更糟的时候，伤害他人还能够是错误的吗？如果我们本可以请求这些人同意却没能这么做的时候，这可能是错误的。由于没有请求这些人的同意，我们侵害了他们的自主权。但这不能够是对我们选择冒险政策的诘难。由于我们本不能与生活在距今许多世纪之后的人们进行可能的交流，所以，我们就不可能请求他们同意。

当我们不能够请求某个人的同意的时候，取而代之，我们应当问这个人以后是否会痛惜我们现在正在做的事。那些后来丧生的人们会痛惜我们选择冒险政策吗？且允许我们假设这些人知道那些事实中的一切。从早年他们就知道，由于核辐射的泄漏，他们身患不治之症，将在他们40岁的时候夺去他们的生命。他们也知道，如果我们本来选择的是安全政策的话，他们就根本不会出生。这些人会痛惜自己英年早逝这个事实。但是，既然他们的生活是值得过的，他们不会痛惜他们出生这个事实。他们因而不会痛惜我们选择冒险政策。

当我们既知道倘若受到伤害的人们对我们的行动了然于心就不会痛惜这个行动，又知道我们的行动对这些人而言不会比我们别的行动的时候更糟，伤害他人能够是错误的吗？我们如何可能知道，尽管我们在伤害某个人，我们的行动对这个人而言将不是更糟的？至少有两种情况：

（1）尽管我们在伤害某个人，但我们可能还知道我们是在赋予这个人某种予以充分补偿的益处。除非该益处明显地在权重上重于该伤害，我们就不会知道此点。但是，如果情况就是如此，我们在做的对这个人而言将更好一些。在此种情况下，如果我们也没在侵害这个人的自主权，对我们的行动就可能没有什么诘难。当我们既知道这个人将没有什么痛惜之处，又知道我们的行动对这个人而言明显地更好一些的时候，对我们伤害某个人可能没有什么诘难。在英国的法律中，外科手术一度被认定为正当的严重身体伤害。正如我在第25节中所论述的，我们应当修正“伤害”一词的日常用法。如果我们在做的是

对某个他人而言将不是更糟，或者对这个人来说甚至更好的话，在某种道德相关的意义上我们并不是在伤害这个人。

如果我们假定使其存在能够带来益处，我们所作的冒险政策选择，在其对那些丧生的人们的影响方面，就像外科手术医生的情况一样。尽管我们的选择造成这些人丧生，但鉴于它也造成他们存在，拥有一种值得过的生活，它赋予他们在权重上超过这个伤害的一种益处。这提示着，我们的选择伤害到这些人这一点不能够作为对我们的选择的诘难。

我们可能代之以假定使其存在并不带来益处。照此假定，我们选择冒险政策并不赋予那些死于该政策之手的人们某种予以充分补偿的益处。我们的选择对这些人们来说不是更好。对他们而言只是不更糟。

（2）还有一种情况，其中我们能够知道，尽管我们在“伤害”一词的日常意义上正在伤害某个人，但是对这个人而言这并非更糟。这是一些涉及到多因素决定的事例。在这些事例中我们知道，如果我们并不伤害某个人，这个人在别的方面至少将受到同样程度的伤害。假设某个人身陷失事的船只当中，就要被烧死了。这个人请求我们开枪打死他，以便不至于痛苦地死去。如果我们杀死这个人，我们并不是在某种道德相关的意义上伤害这个人。

这样的事例并不能够表明，对我们选择冒险政策没有什么诘难，因为它并不相关地相似。如果那场灾难并未发生，那些罹难的人们本会再活许多年。我们选择冒险政策对这些人而言之所以并非更糟，是因为另有相当不同的理由。

会有这样一种情况，在其中我们杀死某个存在着的人，其时知道我们在选择冒险政策时所知道的吗？我们必须知道（a）这个人将获悉但并不痛惜我们已经做了某种要造成他丧生的事。而且我们必须知道（b），尽管这个人否则会正常地再活许多年，造成他丧生对他而言将既不更好，也不更糟。如果假定，我们正在做导致那些被灾难夺去

了生命的人的存在原因中的一个必须的部分的时候，我们并不能正在使这些人受益，那么（b）就是我们对选择冒险政策所带来的后果的了解。

假设我们杀死某个存在的人，要不然他本会再正常活上许多年。在这样的一个情况下，我们不可能会知道（b）为真。即便再活这么多年对这个人来说既不更好，也不更糟，这也根本不会被预言到。不能有这样一种情况，其中我们杀死某个存在的人，其时知道我们在选择冒险政策时所知道的。一个相关相似的事例必定涉及到造成某个人被杀死，如果我们本来不这样做的话，这个人根本不会存在。

请比较这么两个事例：

> 珍妮的选择。珍妮患有一种先天性的疾病，此病将会在她大约40岁的时候毫无痛苦地夺去她的生命。这种疾病在夺去人的性命之前没有什么影响。珍妮知道，如果她生一个孩子的话，这个孩子将罹患同样的疾病。假设她还能够假定下述情况。像她本人一样，她的孩子将有值得过的生活。不存在什么需要领养却尚未被领养的孩子。鉴于这个事例发生时（大概在某个未来的世纪）的世界人口，如果珍妮生一个孩子，对其他人而言将并非更糟。而且，如果她不生这个孩子的话，她将不能够抚养一个孩子。她不能够说服其他人多生一个她本可以抚养的孩子。（这些假定给我们造成那个相关的问题。）由于知晓这些事实，珍妮选择了生一个孩子。

> 露丝的选择。露丝的病症就像珍妮的病症，但有一个例外。她的先天性疾病有一点不像珍妮的疾病，就是只要男性婴儿的命。如果露丝支付新兴的体外受精技术费用的话，她肯定会生一个女儿，此病不会要女儿的命。她决定省下这笔费用，冒险碰运气。不幸的是，她生了一个儿子，他所遗传到的疾病将在大约40岁的时候要他的命。

对珍妮的选择存在一种道德诘难吗？鉴于该事例中的那些假定，这个诘难会不得不诉诸珍妮的孩子所受到的影响。她的选择对这个孩子而言并非更糟糕。对她的选择存在一种诉诸这个孩子权利的诘难吗？假设我们认为，每一个人都有享尽天年的权利。珍妮知道，如果她生养一个孩子，他享尽天年的权利不可能会得到实现。这可能蕴含着，珍妮并未侵犯这个权利。但是这个诘难可以重新表述。可以这样

说：“造成一个享有不能实现的权利的人的存在是错误的。这就是为什么珍妮的行动错误的原因。”

这是一个正确的诘难吗？如果我是珍妮的孩子，我的观点就会像那个写信给《泰晤士报》的男子的观点。我会痛惜我将英年早逝这个事实。但是，既然我的生活是值得过的，我不会痛惜我的母亲给予我的存在。而且我会否认，由于她生下我，她的行动是错误的。如果我被告知说，因为它造成我带着不能实现的权利而存在，所以它是错误的，我会放弃这个权利。

如果珍妮的孩子放弃了他的这项权利，或许会从根本上削弱对她的选择的这个诘难。但是，尽管我会放弃这项权利，但是我不能肯定，在所有这样的情况下，这就是这样的一个孩子会做的。如果珍妮的孩子并不放弃他的这项权利，诉诸这个权利恐怕就提供了某种对她的选择的诘难。

现在转到露丝的选择。对这个选择显然存在一个大得多的诘难。这是因为露丝具有一个不同的替换选择。如果珍妮不生一个孩子的话，她将不能抚养孩子；而且将少一个活着的生命。露丝的替换选择是为那项新技术支付费用，这项新技术将给她一个不同的孩子，她的疾病将不会要这个孩子的命。她选择省下这笔开支，其时知道她的孩子将被疾病夺去生命的机会是一半。

即便对珍妮的选择存在一种诘难，但对露丝的选择存在更大的诘难。这个诘难不能够只诉诸露丝事实上的孩子所受到的那些影响，因为这些影响就像珍妮的选择对珍妮的孩子的那些影响一样。对露丝的选择的诘难必须部分诉诸它对一个不同的孩子的可能影响；通过支付那项新技术的费用，她本会有一个不同的孩子。诉诸这个影响并不是诉诸任何一个人的权利。

现在回到我们的冒险政策选择。如果我们选择这个政策，可能造成那些将在灾难中丧生的人们的存在。我们知道，对这些人们而言，我们的选择并不会更糟。但是，如果对珍妮之选择的诘难有力量的

话，这个诘难就会适用于我们的选择。通过选择冒险政策，我们可能导致这样的一些人们的存在，而他们尽享天年的权利却不能得到实现。

诉诸这些人的权利可能提供某种对我们的选择的诘难。但是它并不能够提供整个诘难。我们的选择在一个方面与珍妮的选择不一样。她的替代选择是不生孩子。我们的替代选择就像露丝的选择一样。如果我们已经选择的是安全政策的话，我们就会有不同的后裔了，他们之中也就没有什么人会被泄漏的辐射夺去生命了。

对露丝之选择的诘难不能够只是诉诸她的孩子尽享天年的权利。对我们的冒险政策选择的诘难因而亦然。这个诘难必须部分诉诸对那些可能的人们的影响，如果我们本来做出了不同的选择的话，这些人本会存在。就像以前，诉诸权利不能够完全解决非同一性问题。我们还必须诉诸一个像Q那样的主张，这一主张对两组不同的可能生活进行比较。对我们的冒险政策选择的诘难因而亦然。这个诘难必须部分诉诸对那些可能的人们的影响，如果我们本来作出了不同的选择的话，这些人本会存在。如前所述，诉诸权利不能够完全解决非同一性问题。我们还必须诉诸一个像Q那样的主张，这一主张对两组不同的可能生活进行比较。

它可能受到这样的诘难："当露丝孕育她的孩子的时候，这个孩子就遗传到了使之不能享尽天年的疾病。因为这个孩子的疾病是以这种方式遗传到的，不能够说露丝的选择杀死了她的孩子。如果我们选择冒险政策，因果联系就更不紧密。因为这种联系更不紧密，我们的选择杀死了那些后来死于核泄漏辐射的人们。我们杀死了这些人就是对我们的选择的充分诘难。"

这个诘难我发现是值得怀疑的。凭什么由于因果联系较不紧密我们的选择就受到大得多的诘难？这个诘难在它有关我们"杀死"一词的日常用法方面所持的主张可能是正确的。但是，正如我在第25节中所

论述的，这种用法在道德上是不相关的。既然那个论证可能不够令人信服，我添加

> 冒险的不孕疗法。如果安不接受某种治疗，就不可能生养孩子。如果她接受这种治疗，她会生养一个儿子，他将是健康的。但是存在着一种危险，就是这种治疗可能令她患上一种罕见的疾病。这种疾病具有下面一些特征。它是无法检测出来的，而且并不伤及妇女，但是会传染最亲近的一些亲属。下面的情况因而为真。如果安接受这种治疗并生下一个健康的儿子，那么她以后将会传染给她的儿子这种疾病的几率占一半，而这种疾病在他大约40岁的时候将夺去他的生命。安选择了接受这种治疗，而且后来把这种致命的疾病传染给了她的儿子。

照上述的那个诘难，对安的选择存在一种强烈的诘难，而这个诘难并不适用于露丝的选择。因为因果联系较不紧密，安的选择杀死了她的儿子。而且她明知她的选择会有这种结果的几率占到一半。露丝知道，她的孩子将于大约40岁的时候死去，几率相同。但是，因为因果联系是如此紧密，并非她的选择杀死她的儿子。根据这个诘难，这个差异具有极大的道德相关性。

这没有令人信服之处。露丝和安两个人都知道，如果她们以某种方式行动，她们将要生下的儿子在大约40岁的时候会被某种疾病夺去生命的几率占到一半。因果说则与之不同。但是这并不使安的选择在道德上更糟。我认为，这个例子表明，我们应当拒斥最后这个诘难。

诘难者或许说：“我否认安通过选择接受那种冒险疗法而杀死了她的儿子。”但是，如果诘难者如此否认，他就不能够主张，我们通过选择冒险政策而杀死遥远的未来中的一些人。因果联系呈现同样的形式。每一选择都产生一种副作用，这种副作用后来杀死那些其存在归因于这一选择的人们。

如果这个诘难不成立了，就像我所认为的，我早先的主张就得到了正当性证明。如果我们选择冒险政策，我们的选择像露丝的选择而不是像珍妮的选择，在道德上意义重大。如果我们本来作了别的选择的话，本来会生活着不会那样丧生的一些不同的人，在道德上意义重

大。既然如此，对我们的选择的诘难就不能够只诉诸那些后来事实上生活着的人的权利。还必须诉诸像Q一样的一个主张，这一主张对两组不同的可能生活进行比较。正如我前面所主张的，诉诸权利不能够完全解决非同一性问题。

127.结论

我现在对我所主张的加以总结。的确，就每一个人来说，如果当他成胎时的1个月之内尚未成胎的话，他根本就永远不会存在。因为此点为真，所以我们能够轻易地影响到未来的人的同一性，或者说，影响到将于以后生活着的那些人到底是谁。如果两个社会政策之间的选择将影响到长达大约一个世纪的生活水准和生活质量，那么它将影响到以后在我们的共同体中所过的所有生活的那些细节。结果，那些于以后生活的那些人中的一些人的存在，将归因于我们选择了这两个政策中的一个政策。在经过一个或者两个世纪之后，这对我们的共同体中的每一个人来说都如此。

这一事实产生一个问题。这两个政策中的一个政策在更遥远的未来可能造成生活质量的大幅度下降。这会是我称为消耗的那项政策的后果。这个后果是糟糕的，而且提供了不要选择消耗的一个道德理由。但是，由于刚刚提到的那个事实，我们选择消耗不会对什么人来说将更糟。一些人认为，如果一个选择不会对什么人来说是糟糕的，那么这个选择就不会有糟糕的影响。消耗事例表明，我们必须拒斥这个观点。冒险政策事例更加强有力地说明了这一点。选择这项政策所带来的一个影响是一场杀死数以千计人的灾难。这个后果显然是糟糕的，即使我们的选择没有对什么人来说更糟。

既然这两个选择不会对什么人来说将更糟，我们需要解释为什么我们具有不要作这些选择的道德理由。这个问题之所以产生，是因为在不同的结果中会有不同的一些人存在。我因而称之为非同一性问题。

我问了我们是否能够凭借诉诸人们的权利来解决这个问题。我论证说，即便是在那个冒险政策事例中，对我们的选择的诘难也不能够只诉诸人们的权利。该诘难必须部分诉诸像Q这样的主张，这一主张对两组不同的可能生活进行比较。而且，在解释对我们选择更小的消耗的诘难方面，我们不能够令人信服地诉诸权利。即便在生活质量大幅度滑坡之后，那些将要生活着的人们也将比我们现在的境况好得多。这些人们无权享有那个甚至更高的生活质量，倘若我们本来选择了保护的话，那个甚至更高一些的生活质量会由一些与他们不同的人享有。如果我们设想离开非同一性问题，对我们的选择的诘难会诉诸我们的善行原则。为了解决非同一性问题，我们必须修正这个原则。

一个修正过的原则是Q主张，即相同数目质量主张。根据Q，如果在无论这两个结果的哪一个结果中，都存在相同数目的人，如果那些生活着的人比那些本会生活着的人们境况更糟，或者生活质量更低，那就会更糟一些。我们需要一个更广泛的原则，来涵盖那些不同的结果中会有不同数目的人的事例。这个所需的原则我称之为X理论。只有X理论将充分解决非同一性问题。

非同一性这个事实造成一种道德差异吗？当我们看清我们的消耗选择不会对什么人来说将是更糟的时候，我们可能认为对我们的选择所存在的诘难较小。但是我认为，诘难同样强烈。而且当我比较那两个医疗方案的时候，我具有相似的信念。这个信念我称之为无差异观点。尽管我知道有一些人并不接受这个观点，但是我知道接受这个观点的人更多。如果我们接受无差异观点，而且认为使某人存在不能带来益处，这具有广泛的理论性蕴含。我们能够预见到，X理论将不会呈现为一种对人产生影响的形式。有关善行的最佳理论将不诉诸对那些受到我们行为影响的人们来说何为好何为糟。

在下文中，我将努力找到X理论。正如我所主张的，这一尝试将引发一些令人困惑的问题。

---

[1]Raverat。

[2]参见Kripke，Bogen，Forbes（1）。

[3]参见Kripke中对这些观点的讨论和所提到的参阅。

[4]一些人主张，不存在什么区别性的实质属性。这蕴含着，对于这些人中的各人而言，假如他的父母本来从未婚配的话，他是否本来永远不会存在这个问题是一个空洞的问题。尽管我认为有关我们的同一性会有空洞的问题，但是我怀疑，经过反思之后这些人是否会认为这个问题是空洞的。（在同卵双胞胎这一特殊事例中，此处有一些空洞的问题。我希望另寻别处来讨论这一点。）

[5]这个主体与历时的人格同一性之间的关系是什么？根据起源的观点，每一个人从一个受精卵成长起来这一点是他的一个实质属性。这个观点会与物理准则结合起来。按照这个准则的那个最令人信服的版本，每一个人的一个实质属性是他有足够的他的特定大脑来支持完全的意识生活。或许有人主张说，成长自一个特定的受精卵这一点是任何特定大脑的一个实质属性。

物理准则无需与起源的观点相结合。这一准则的信奉者可能接受反向变量观。按照这个观点，托尔斯泰或许本来有一个不同的起源。在这个想象出的不同的可能历史中，物理准则本来会得到实现。相似地，我们会接受起源的观点，而拒斥物理准则。我们可能把起源的观点与心理准则的那些广义版本结合起来。那么我们会认为，我成长自一对特定的细胞并且从而开始生活在一个特定的躯体中这一点对于我而言是实质性的。但是我们也会认为我的生活会在一个不同的躯体中继续。例如，如果我被电子传输，这一点就会发生。

下面考虑心理准则，这个准则诉诸心理的连续性。那些相信这个准则的人们或许同意，我的生活本来会不同地进行。而且在实际生活中的我与这个不同的可能生活中的我自己之间或许有非常小的心理连续性。假设我的父母在我3岁的时候把我带到意大利，而且我变成了一个意大利人。那么在现在的我与我本来现在会是的我自己之间会有下述关系。在我的实际生活和这个不同的可能生活这两者之内，我现在都会与我3岁时候我的实际生活中的我自己具有心理上的连续。这个关系会是微弱的。在现在的我与3岁时的我自己之间几乎没有多少直接的心理联系。如果我们比较我的实际生活与这个不同的可能生活，那么，这两个生活在成年时期甚至可能不包括哪怕一个共同的记忆。

心理准则的一些信奉者可能主张说，这些并不是同一个特定的人的两种可能的生活——那个3岁的时候本来会到意大利的孩子本来就不会是我，而是不同的一个人。这个主张修正了我们对于个人同一性的通常观点。

这个修正在技术上既是困难的又是不必要的。可以更简单的将这一要点加以表达。假如我3岁的时候去了意大利的话，我的生活本来会非常地不同。而且我们可能相信这个事实具有各种各样的实践上的和道德上的蕴含。但是这个信念无需通过否定实际生活中的我与这个不同的可能生活中的我自己之间的同一性来表达。我们可以承认这个关系是同一性。我们能够通过这样的主张来表明我们的要点，就是，当我们在比较这样的非常不同的可能生活的时候，个人同一性这个事实并不具有其通常的重要性。参见Adams（3）。

[6]出于布赖恩·巴里的《正义的情状和未来世代》所给出的理由，载Sikora and Barry。

[7]参见《家庭场合中的青少年怀孕：对于政策决策的蕴含》，西奥多拉·乌姆斯（娘家姓帕菲特）编，坦普尔大学出版社，费城，1981年。

[8]这个难题被Kavka称作未来个体悖论。参见Kavka（4）。

[9]我沿袭Adams（3）。

[10]参见T.Schwartz，《对于后裔的义务》，载Sikora and Barry。

[11]该诘难的这个形式是Tooley提出的。

[12]参见B·巴里，"能源政策中的代际正义"，载Maclean and Brown；另参见Barry（2）。

[13]至于进一步的讨论，参见J·伍德沃德，《非同一性难题》，载《伦理学》，1986年7月。

[14]J·麦克马汉向我提出，如果这个残疾对这些人的生活的性质影响很大，那么假如某个终身残疾的人生下来是正常的话他本来处境会更好这一点可能并不是清楚的。一些人可能怀疑在那个相关意义上这两个非常不同的生活是否本来为同一人所过活。而且Adams（3）提出，即便这样的一个人本来处境会更好，这也无需蕴含他对自己的残疾并不遗憾这一点会是非理性的。如果我们接受两者中的任何一个主张，那么该事例就并不是我们所需要的。我们能够通过假设该残疾只是在成年时才影响他们而避免这些问题。例如，该残疾可能是不能生育。

[15]看起来可能有一个例外。如果我的生活是值得过的，杀死我就是往更糟的方面影响我，但是它造成我处境更糟，或者造成我具有一种更低的生活质量吗？正如我对这个短语的用法，我的确具有"一种更低的生活质量"。倘若我的生活比它本来会有的较不好一些，或者倘若在我的生活中所发生的对于我更糟，那么这一点就为真。当我过着值得过的生活而遭到杀害的时候这两者皆真。

[16]例子之一是Scanlon（3）中所提出的那个令人信服的理论。Scanlon论证说，对于道德动机的最佳说明不是功利主义者们所提供的说明，他们诉诸的是普遍的博爱。相反，我们根本的道德动机是"立足他人不能有理地予以拒斥的根据而能够证明一个人给他人所施加的行动是正当的这一愿望"。Scanlon基于这个主张勾勒出一个吸引人的道德理论。按照这一理论，如果一个行动将以一种无法证明正当的方式影响到某个人——如果有某个其抱怨不能得到回答的抱怨者存在，那么这个行动就是错的。按照这一理论，道德的框架实质上是对人产生影响的。不幸的是，当我们选择一个像更大的消耗政策的时候，将没有什么抱怨者。如果我们认为这并不产生什么道德上的影响，鉴于对于我们所做选择的那个诘难一样强，那么我们认为将没有什么抱怨者这一点是不相关的。Scanlon理论的根本原则在按照我们的观点看来不应当作出什么区分之处作出了区分。Scanlon的理论因而需要修正。
相似的说法适用于许多别的理论。于是Brandt（2）提出，我们应当赋予"道德上是错的"这个短语这样一种描述性意义，这种意义"会遭到所有完全合理的人们会倾向支持的任何道德法则的禁止，这个道德法则优先于一切其他道德法则或者之外根本没有什么其他道德法则，为的是行动者的那个社会，如果他们期望在那个社会中度过终生的话"（第194页）。看起来很有可能的是，按照这个所选的法则，如果没有什么抱怨者的话，那么一个行动就不会是错的。相似的说法适用于格特、纳维森《道德和功利》，适用于G·R·格赖斯《道德判断的根据》，而且可能适用于Mackie（2）；Richards，Harman，Gauthier（4）；Rawls，等等。

第十七章

令人讨厌的结论

应当有多少人？会有人口过剩——太多的人吗？对这些问题我们需要一个答案，这个答案也解决非同一性问题。

我后面将问到底应当有多少人。在一个完全的道德理论中，我们不能够回避这个令人敬畏的问题。而且我们的回答可能具有一些实际蕴含。例如，它可能影响到我们有关核武器的观点。

在下文的绝大多数篇幅中，我讨论一个较小的问题。在某个特定阶段，某个国家或者世界应当有多少人？什么时候生活着的人会过多？

128.更多的人生存就更好吗？

考虑一下

> 幸福的孩子。一对夫妻正在试图决定是否要生另一个孩子。他们能够假定，如果他们又生了一个的话，他们会爱这个孩子，而且他的生活会相当值得过。鉴于这个事例发生时（大概未来的某个世纪）的世界人口，这对夫妻能够假定，再生一个孩子总的来说对其他人来说不会更糟。当他们考虑对他们自己的影响的时候，他们有多个再生另一个孩子的理由；他们同样有多个相反的理由，诸如对他们生涯的影响等。像许多别的人一样，这对夫妻不能够在这两组相互冲突的理由之间作出决定。他们认为，如果他们生了这个孩子的话，这对他们而言既不更好也不更糟。

这对夫妻有生这个孩子的道德理由吗？如果多生一个的生活会更好一些吗？一些人给出的回答是肯定的。如果这对夫妻有生养这个孩子的一个道德理由，而且他们不能够在他们其他的那些理由之间作出决定，他们的这个道德理由可能使天平出现倾斜。大概他们应该生这个孩子。

其他人持一种与之不同的观点。他们认为，没有什么道德理由来生这个孩子。照他们的观点，多生一个的生活不会更好。

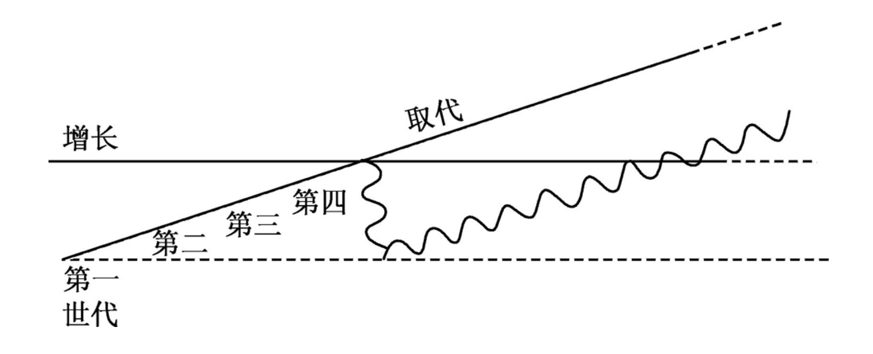
129.人口增长对现存的人的影响

我所说的那对夫妻假定，一个额外的孩子的存在对其他人们而言总体上说不会更糟。在许多国家中，在许多时期，一直如此。但是在其他一些时期，并非一直如此。在这些时期，如果人再多的话，人们的境况本来会更糟。现在的许多国家就是如此。在这些国家中，如果人口增长，生活质量将比人口不增长的话要低一些。这些就是我将要予以讨论的事例。

在这些事例中，人口增长会影响到现存的人。当人口增长降低生活质量的时候，我们可以说人口增长必定有悖于现存的人的利益。但情况并非总是如此。甚至当人口增长降低生活质量的时候，人口增长对现存的人们而言也能够更好一些。

在某个国家的某个阶段，下述情况能够为真。如果人口按照一定的速率增长，这将产生暂时好的后果，累积性的糟糕后果。那些糟糕后果可能是人均可支配资源的持续走低。暂时的好后果或许及于这个国家的经济运行。尽管快速人口增长可能对经济而言是糟糕的，但是缓慢的人口增长可能比人口一点也不增长对经济而言要更好。对某些经济体制而言这能够成立的技术性理由多种多样。一个冷僻的类比是这样一个事实，当我们在弯道上驾车的时候，如果我们加速行车会更容易掌握方向盘。而且，人口增长可能还具有一些暂时性的非经济的好后果。这些或许简单如大家庭总是稍微幸福一些，抑或许多人偏爱有更多的孩子。

罗伯逊写道，如果人口增长的某个速率会有这两类后果，那么，"一个总是增长却永远不变大的人口可能会是……最为合适的"[1]。这样的理想只有在《爱丽丝漫游奇境记》中的梦幻世界里才会实现。在现实世界中，可供选择的余地能够用下图来表示：



我称之为下行的自动楼梯事例。一条线表示保持一个稳定的人口所导致的生活质量，或者我所称的取代。那条间断的虚线表示某个人口增长速率的累积性糟糕后果。这条线表示如果没有暂时性的好后果的话，生活质量会是什么样子。另一条连续不间断的实线表示这个增长速率的综合后果。这是在假若有暂时性的好后果的情况下生活质量会是的样子。

正如本图所示，如果出现的是增长而不是取代，对前三个世代的人们而言生活质量将更高，但是随后将越来越低。许多人认为，如果人口增长降低了生活质量的话，会是糟糕的。对这些人而言，下行的自动楼梯事例尤其令人压抑。看来很可能，实际上的后果会是增长。如果大多数夫妻本人有两个以上的孩子，对他们而言可能更好；而这可能是大多数夫妻想要做的。况且，既然增长对现存的人们和接下去两个世代中的人来说会更好，那么共同体会决定某项造成增长变为取代的政策就不大可能。

或许有人说："情况并非如此。最佳的政策对前三代人来说会是增长，取代政策用于其后。一旦增长就会产生一种较低的生活质量，共同体就应当把政策调到取代。在此点增长对现存的人来说不再会更好一些了。既然如此，我们可以期待共同体调整到取代政策。"

这大错特错了。考虑一下第四代人所面临的选择余地。如果对前三代人而言一直采取增长，生活质量为什么仍然像倘若本来采取取代的话会有的那样高？只是因为增长的暂时性好后果。如果第四代人调整到取代，将会丧失这些好的后果。其生活质量会垂直降到与那条虚

线相平的程度。而且如果继续采取取代政策，对接下去的三个世代的人来说，生活质量将比倘若本来有人口增长的话会有的要低。这两种后果由波状线表示。正如波状线所表示的，第四代人在那些较低的水准点上具有与第一代人所具有的相同的选择余地。每一代人都有这些选择余地。

只要增长具有这两个后果——累积性的糟糕后果和暂时性的好后果，这些就依然是些可供选择的余地。尽管此点为真，但是相较于取代，增长对现存的人和下两个世纪的人总都更好一些。因而非常可能的是，每一代人都将选择增长。结果，生活质量持续滑落。如果我们认为这是一个糟糕的后果，正如我所主张的，下行的自动楼梯事例是个特别令人压抑的事例。它是一个代际的囚犯的两难困境，是那种相关的人在其中最不大可能获得解决方案的两难境地。[2]

我们可以假定，在判断我们的行动后果方面，考虑到所有那些将会生活的人们的利益就足够了。在下行的自动楼梯事例中，人口的某种增长速率将总是符合现存的人们及其子孙的利益。尽管它造成生活质量连续下滑，这个增长速率将不会有悖于那些晚于三代人以上才过活的人的利益。这是因为，按第123节中所解释的那种方式，这些人的存在归因于这个增长速率。如果我们只诉诸所有那些将会生活的人们的利益，我们必定主张，尽管出现生活质量的持续走低，如果存在这个增长速率情况会更好一些。如果我们想避免这样的结论，这就又涉及到一种情况，其中我们必须诉诸与之不同的一种原则，一种并非对人产生影响的、亦非与人们利益有关的原则。

对那些探究生活质量滑坡的人们而言，这是最令人沮丧的情况。幸运的是，这只是数个可能的情形中的一种而已。当人口增长降低生活质量的时候，人口增长对现存的人们而言能够更糟的情况有二。

在某些共同体中，如果大多数夫妻本身有两个以上的子女的话，对他们而言将更糟。这些是其中最不大可能出现人口增长的那些情况。

在另一些共同体中，就像我在第三章中所解释过的，大多数夫妻面对一种囚犯的两难困境。在这些共同体中，如果许多人不管别人怎么做，他或者她都有两个以上的孩子，对这许多人中的每一个人而言都将更好。但是，如果所有的人都有两个以上的孩子，对每一个人而言就比在没有任何人这样做的情况下更糟。如果这些人最终看清了这个实情的话，他们或许会取得我所称的一种政治解决方案。尽管每一个人都会倾向于生养更多的子女，但是每一个人或许也倾向于：与其人人都生养更多的子女不如没有任何人生更多的子女。一种旨在终止人口增长的奖惩制度或许被民主地采纳。即使这种制度是非民主地强加于人的，这样的一种制度或许会得到所有这些人的欢迎。生育两胎之后实施可逆绝育措施，或许提供了另一种解决方案。这是一种更好的解决方案，因为它不会施加任何惩罚。而且如果他们理解那些事实的话，所有这些人可能会欢迎这个解决方案。

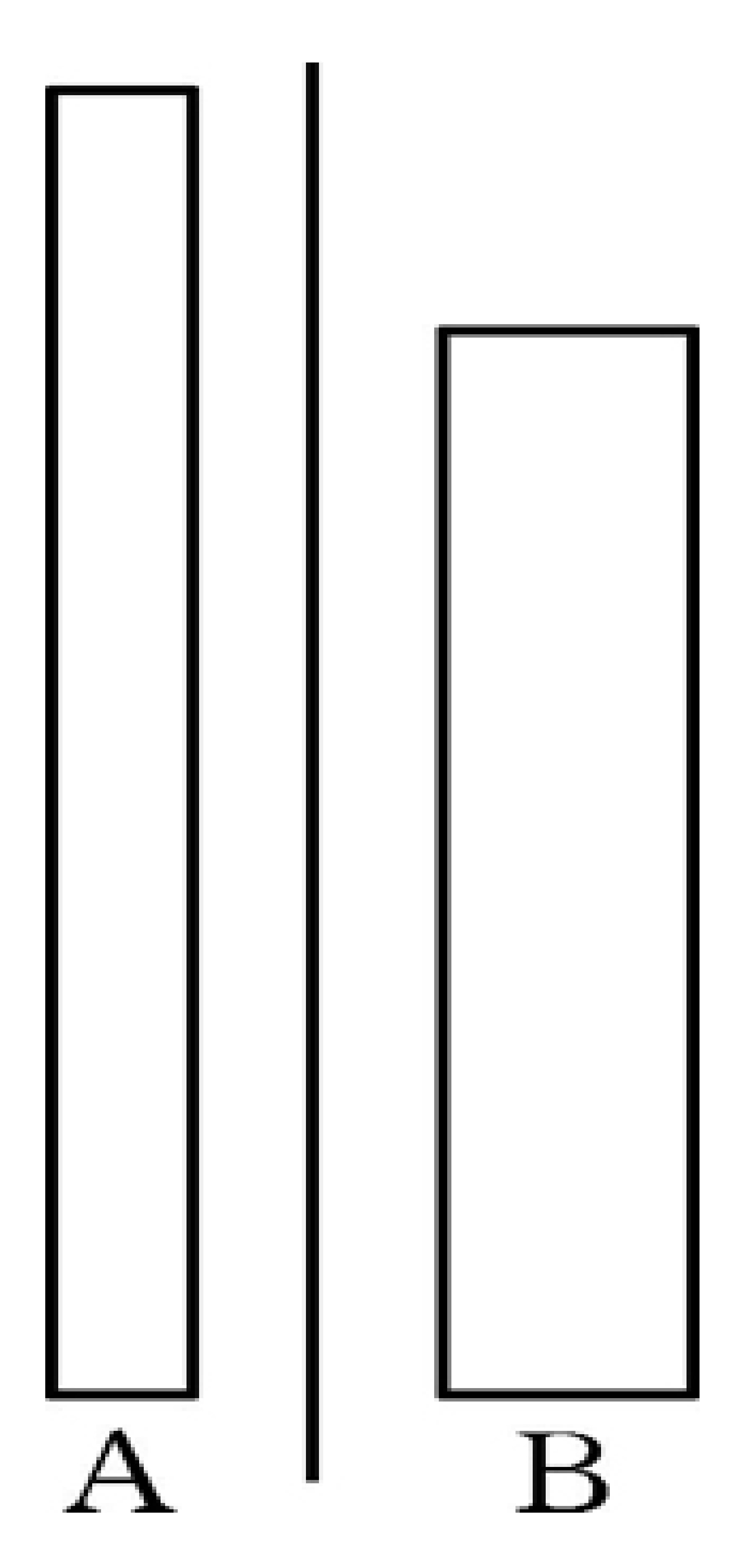
我业已表述了三种情况。而且还有其他一些可能的情形。这些情形或许是这三种情况的混合；但是另一些情形会在别的方面有所不同。

130.人口过剩

当人口增长降低生活质量的时候，对现存的人的后果可能好、可能糟，也可能既不好也不糟。这些影响并不引发新的道德问题。但是，其他一些影响的的确确引起这样的一些问题。

当我们比较不同的人口增长速率在更远的未来所生发的那些结果的时候，这些问题的出现就再清楚不过了。如果有更快一些的增长，后来将有更多的人，这些人的境况将更糟。如前所述，“境况更糟”既可以指涉幸福的水准，也可以指涉生活质量，抑或指涉人均资源占有额度。我们应当假定，在我所举的那些事例中，这三者相互协同，同升同降。

让我们比较一下两种人口增长速率在一个或两个世纪之后所带来的结果。正如我所解释的，没有什么人会在这两种结果中都存在。两个这样的结果如下图所示：


A
B

每一方块的宽度表示生活着的人的数目，高度则表示他们的生活质量。借此我意指他们在某个时间阶段中的生活质量。在这样的一个阶段中，人口可能会有某种变化。但是出于简单性起见，我们可以忽略这个事实。同理，我们可以假定，在这些结果中既没有社会的也没有自然的不平等；没有什么人的境况比任何别的人更糟。事实上永远不会如此。但是，如果我们如此设想的话，也不会使我就有关我要问的那些问题所作的推理失真。而且这使我的那些问题呈现出一种更加清晰的形式。

B中的人数是A中人数的两倍，而且都比A中的任何一个人的境况要糟。但是B中的那些人的生活，与A中的那些人相比较而言，只有值得生活程度的一半多的程度。如图所示，这个主张并非假定这些判断原则上会是精确的。我认为只有粗略的或者部分的可比性。我的主张所假定的是，从A中的水准移动到B中的水准会是生活质量上的一种滑落，除非它会比在人们的生活不再值得过之前的另一个类似大的滑落程度大得多。

受两倍人口的拖累，生活质量的降低或许是多方面的。或许会出现住房条件恶化、学校拥挤、污染加剧、自然美景减少，以及较低的人均收入水准。如果这些是生活质量降低的方方面面，我们能够令人信服的假定，它会比生活不再值得过之前的另一个类似滑落程度大得多。

除了不存在不平等之外，这两个结果会是某个国家或者整个人类的真实选择余地，假若两种人口增长速率保持许多年的话。哪一个是更好的结果？我们用“更好”这个短语并不意指这个短语最常见用法上的“道德上的更好”。此含义只适用于人或者行动。但是两个结果中的一个能够在具有道德相关性的另一个意义上更好。在后者这个意义上，如果更少的人患有某种令人致残的疾病，或者里斯本大地震根本没有发生过的话，情况会更好一些。关于涉及到不同的可能人口的那些结果，我们能够清楚地持这样的一些主张。假设在两个这样的结果

中，会存在相同数目的人。如果，在其中一个结果中，人们的境况会糟得多，这明显就是更糟的结果。这个结果不会对什么人来说会更糟。但是，正如我所论证的，这并不表明这个结果不可能是更糟的。

回到A和B。哪一个结果会更好？B中的人们境况更糟，这显然是糟糕的。这能够在道德权重上被更多的人生活着这个事实所超越吗？

假设我们认为在幸福的孩子事例中，我所说的那对夫妻没有什么生这个孩子的道德理由。那么我们可能相信，如果人们境况更糟，这不能够在道德权重上被生活着的人的数目的增加而压倒。那些相信这一点的人们时常诉诸

> 非人称的平均原则：如果在其他事情上是同等的，那么最好的结果就是其中人均生活而言过得最好的生活。

一些经济学家借助定义来使这个原则成立。[3]我之所以称这个原则是非人称的，是因为它并不是对人产生影响的：它不是关于对那些受到我们行动影响的人们而言会是好的还是糟的。这个原则并不假定，如果人们存在，并且具有一种值得过的生活的话，这些人从而是得到益处的。

这个原则的快乐主义版本认为

> 如果在其他事情同等的情况下，那么最好的结果就是其中每个生活过的人享受最大幸福均值的那个结果。

我是以时间中性的形式陈述这些版本的。一些人陈述这个平均原则的方式则只指称那些我们已经行动之后才活着的人们。在这种形式中，该原则荒谬地蕴含着，现在活着的人们中间除了最迷狂出神者之外所有的人都被杀掉会更好一些。照平均原则的时间中性版本，如果某个具有值得过的生活的人早逝，这造成人们的生活平均来说变得更糟了。

下面假设我们认为在“幸福的孩子事例”中，我所说的那对夫妻有一个生这个孩子的道德理由。我们认为，如果一个额外的生命过着值得过的生活，这本身总是更好一些。如果这是我们所相信的，那么就我所说的两个结果而言，主张B或许比A更好一些就是再自然不过的事了。生活质量方面的损失或许在权重上被所生活过的人的数目方面的一个充分的获益而超过。如果我们作这样的主张，我们必定问：“一个充分的获益会是什么？”

如果我们是快乐主义者，我们能够更为精确地陈述这些问题。我们问

(1)“如果在两个结果的一个结果中，生活着的人们较少不幸福，这能够在道德权重上被幸福的量的一个充分的增加压倒吗？”

如果人们较少幸福，他们就具有一种较低的生活质量。如果我们对问题（1）的回答是肯定的，那么我们必定问

(2)“质量和数量的相对价值是什么？”

非人称的总量原则的快乐主义版本提供了一种回答。这个回答说

如果在其他事情同等的情况下，最好的结果是那个会有最大量的幸福——减去苦恼后幸福的最大净值——的结果。

照这个原则，B会比A更好，因为B中会有更大的幸福量。尽管每一个B都比A的幸福少，但是他们的每一个生活都包含着相应幸福量的一半以上。既然有多达两倍的B，他们合起来具有比A多的幸福量。（两个超过半瓶满的瓶子所容纳的量超过一个满瓶所容纳的量。）

下面假设我们不是快乐主义者。我们所认为的道德上重要的东西不是幸福，而是生活质量。我们能够问相同的一些问题；但是我们必

须使用一个不为人所熟悉的短语。当我们比较质量和数量的价值的时候，什么是相关的量？我们可能说是“值得过的那些生活量”。但这是错误的，因为它忽视了这些额外生活的质量，或者说，忽视了在多大程度上它们是值得过的。相关的量，像幸福总量一样，必定是这些生活的数目和它们的质量这两者的一个函数。为了表述相关量，我提出“使生活值得过的任何一种东西的量”这个短语。

重新考虑A和B。快乐主义者会主张说：“尽管每一个B都比A少些幸福，他们合起来却有更多的幸福。”我们会相似地主张说：“与 A比较下来，较之 A人而言，每个B人所具有的无论什么使生活值得过的东西都要少一些。但是B中的每一个生活都具有A中的每一个生活所值得过的那种程度的一半以上。既然有两倍的B人，那么他们合起来所具有的使生活值得过的任何一种东西的量要更多一些。”

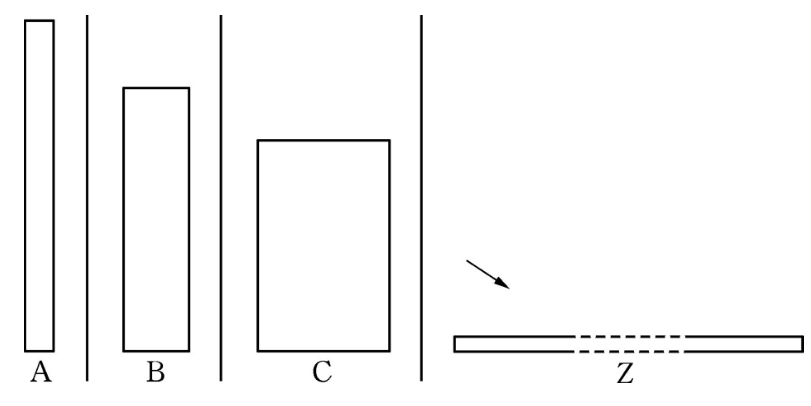
我现在能够陈述非快乐主义的

> 非人称的总量原则：如果在其他事情同等的情况下，最好的结果是其中使生活值得过的任何一种东西的量最大的那个结果。

如果我们认为B会比A糟糕，我们必定拒斥这个原则。

131.令人讨厌的结论

下面考虑一幅更大的图表。



按照非人称的总量原则，恰如B会比A更好，C会比B更好。而Z或许是最好的。 Z是某个庞大的人口，其成员的生活与那个生活不再值得过的水准相比高不了多少。一种生活之所以会是这个样子，要么是因为该人口有足够多的迷狂出神的人，使该人口中的苦恼看起来值得忍受；要么是因为，该人口千篇一律地劣质。让我们设想Z中的生活属于更为单调的第二种情况。在这些生活的每一个生活中，幸福微乎其微。但是，如果数目足够庞大，其结果却是幸福总量最大。同样，Z会是其中使生活值得过的任何一种东西的量最大的一个结果。（积少成多，只在瓶底残留一点牛奶的一堆瓶子或许盛有最大量的牛奶。）

让我们下面假定，出于我后面将予以提出的某个理由，A会有100亿人口。那么非人称的总量原则蕴含着：

> 令人讨厌的结论：对人人享有很高生活质量的至少多达100亿人口的任何一个可能的人口而言，必定有某个可想象的更大的人口，如果在其他事情同等的情况下，这个更大的人口的生存会更好一些，即使其成员只具有勉强值得过的生活。

正如我对这个结论的命名所提示的，我发觉该结论难以接受。

A和B会是事实上的选择余地。 A和Z却不然。一些人主张，有鉴于此，我们无需试图回避那个令人讨厌的结论。他们可能说："既然这个结论并不适用于任何一个可能的选择，那么可以忽略不计。我们无需在那些不会发生的情况中检验我们的那些原则。"

我区分了两类不可能性：深刻的和技术上的。如果一个想象出的事例要求对自然规律的巨大更改，包括对人的自然性质的更改在内，就是深刻地不可能的。有两个向那些深刻地不可能事例挑战的根据。我们可能不会设想这样的一些事例中涉及到什么。而且一些人会主张说，我们的那些道德原则仅仅在真实的世界中是可接受的。[4]

记得诺齐克所设想的功利怪兽可能有所助益。这是些"从他人所失中的所得远远大于他人所失的"[5]人们。这样的一种想象出的人提供了

对行为功利主义的一个诘难，行为功利主义“好像要求我们都沦为怪兽的美餐，以便增加总的功利”。

正如诺齐克所表述的，这样的一种人是一种深刻的不可能性。世界人口现在有几十亿。如果被剥夺了任何超出避免饥馑的那些定额配给之外的东西，以及其他所有流向诺齐克所想象出的怪兽的那些资源，所有这些人的困境我们可以想象一下。诺齐克让我们假设这个想象出的人是如此幸福，或者具有如此高的质量，以至于正是分配产生出最大幸福总量，或者任何使生活值得过的东西的总量。鉴于数以十亿计的人深陷困境之中，而这个怪兽巨大资源中的一小部分就会轻易解除这种困境，这一点何以成立？这一点要想成立，这个怪兽的生活质量必须比我们所知的任何人高几百万倍。对此我们能够想象吗？想一想你所知道的最幸运的人的生活，试问，为了达到值得过的生活的几百万倍，生活得像什么样子。这样的一种生活和我们的生活之间存在的质量的差距，必定类似于我们的生活与那些勉强有意识的生物——如果它们是有意识的话，诸如柏拉图的“满足的牡蛎”[6]——之间存在的差距。说我们即使在最为朦胧的意义上都不能想象这个功利怪兽的生活，看起来也不失为一个公道的回答。而这一点使该例子的力量成了疑问。行为功利主义者或许说，如果我们真的不能想象这样的一种生活会是什么样子，那么我们可能发现诺齐克的诘难不具说服力。他的“怪兽”看起来就像一个类似神的一种存在物。在这样的一个所想象的存在物在场的情况下，我们与其享有同等权利的信念可能开始动摇——就像我们并不认为低等动物与我们享有同等权利一样。

这个回应具有某种力量。但是，即便一个深刻的不可能性也可能提供对我们的道德原则的一种部分检验。我们不能简单地忽视那些设想的事例。

现在回到我所想象出的Z。这个所设想的人口是另一个功利怪兽。不同之处在于，幸福的更大的量并非来自一个人的生活质量的巨大增加，而是来自所过活的生活数目的巨大增加。而我的功利怪兽，

既非深刻地不可能，亦非不能想象。我们能够想象，某个人的生活勉强值得过的话会是什么样子。而且我们能够想象，许多人具有这样的生活的话会是什么样子。为了想象Z，我们只需想象会有非常多的人。对此我们能够做到。所以这个事例不能被当作一个我们几乎不能理解的事例加以置疑。

事实上我们不会面对A和Z之间的选择。如果资源存储是限定的，我们事实上不能凭借生产出其生活勉强值得过的巨大人口，来生成最大的幸福量，或者任何一种使生活值得过的东西的最大量。[7]但这只是技术上不可能而已。为了假设其是可能的，我们只需添加一些有关资源的本质和可供性的假定就可以了。既然面对A和Z之间的选择只是技术上不可能的，所以这并未削弱其作为对我们的那些原则的一种检验方法的可比之处。不同数目选择引发这样一个问题，即生活质量的所失是否总是会被幸福量的增益或者被任何一种使生活值得过的东西的量的增益在道德权重上压倒。这正是比较A和Z时最为清晰地呈现出的问题。如果我们深信Z比A糟糕，那么我们就有充分的根据抵制其中蕴含Z更好一些的那些原则。我们具有充分的根据抵制非人称的总量原则。

也许有人说："并非如此。这个原则还包括如果在其他事情同等的情况下这个短语。其他事情永远不会是同等的。因此，我们可以忽略那个令人讨厌的结论。"

这是并不令人信服的。别的什么道德原则必然被Z的发生所侵害呢？可能有人主张说，这会违反某个有关世代之间正义的原则。但是这与以最纯粹的形式出现的我们的问题不相关。我们在问，如果Z发生的话，是否会比如果A发生的话更好一些。我们可以想象出一种历史，其中只有类似Z的一些结果会发生。那么一来，Z中的人们就不会比任何活着的人的处境更糟。即使我们相信Z会比A 处境更糟，在此也不会是因为Z的发生会涉及到不公正。

还有更重要的一点。再考虑一下非同一性问题。一些人提出，我们可以诉诸人的权利来解决这个问题。但是，正如消耗事例所表明的，情况并非如此。如果我们设想去掉非同一性问题，对我们选择较小消耗的诘难会诉诸我们所说的善行原则。要解决非同一性问题，我们必须修正这个原则。我们必须发现我所称的X理论。

如果我们要避免那个令人讨厌的结论，情况亦然。我们不应通过诉诸涵盖某个不同的道德部分的那些原则来力图避免这个结论。这个结论是内在地令人讨厌的。而且这个结论也为作为善行原则的一个特定版本的非人称的总量原则所蕴含。要避免那个令人讨厌的结论，我们必须努力表明我们应当拒斥这个版本。我们必须找到一个更好的版本：X理论。

---

[1]Robertson，第460页。

[2]值得指出的是，这个事例是具有两个特殊特征的一个时际每人-我们两难。既然它涉及不同的世代，所涉及的那些人们不能够借助交流达成某种政治解决之道，或者某个有条件性的联合协议。而且这是一个特别难以处理的那种包括局外人的两难。
考虑一下那个剧场两难（auditoriumdilemma）。如果第一排起立，那么这将会改善其对舞台上引人入胜的场面的观看视野。如果起立得到这个更好的视野是值得的，那么对于第一排而言倘若站立起来将更好一些。但是这会阻挡第二排的视野。这一排会需要起立，以便重新获得所有的人都坐着的时候所拥有的视野。鉴于现在会是在站着，却不会改善其视野，那么这个结果对于第二排而言就会更糟。相似情况适用于所有的其他各排。
这个事例不同于通常的每人-我们两难。有两个行动：A（更利他主义）和E（更自我主义）。在一个通常的两难中，无论他人做什么，对于各人而言如果他做E则更好，但是对于各人而言如果全都做E那么则比倘若全都做A将更糟。在剧场两难中，存在一个虽小却是决定性的差异。对于各排而言如果站着而不是坐着将更好，但是倘若全都站着而不是坐着，那么并非对于所有的那些排而言都将更糟。它对于除了第一排之外的所有的那些排而言将更糟。第一排在这个两难中是局外人。
因为它们包含局外人，这样的一些两难特别难以处理。对于别的人而言更糟的行动模式对于局外人而言则更好。这样一来，假如局外人帮助达成一个政治解决之道或者加入一个条件性的协议的话，那么对于他们而言就会更糟。而且这些局外人所做的可能启动一个恶的连锁反应，这使加入这样的一个协议这一点成为对于人人而言更糟的事。于是，在剧场两难中，如果全都坐着而不是站着，对于第一排而言将更糟。如果这一排加入一个全都站着的协议，那么对于该排而言将因而更糟。该排因而可能站着。一旦第一排站着，对于第二排而言如果加入一个除了第一排之外全都应当坐着的协议则更糟。它可能因而站着。那么对于第三排而言

加入一个除了前两排之外全都应当坐着的协议则将更糟。它可能因而站着。相似的说法适用于每一排。最终的结果可能是所有排全都站着而不是坐着。这对于除了第一排之外的每一排而言全都更糟。第一排——局外人——的在场，在此防止了条件性的联合协议的达成。而且同一连锁反应可能防止了一个政治解决之道的达成。这个特殊的特征使这样的两难更少可能得到化解。
除了微不足道之外，剧场两难并不具有另一个令人压抑的特征。它涉及到同代人。这使它更有可能得到化解。其他那些排或许使用威胁来保持第一排坐着。或许第一排之所以坐着只是因为它预期到来自其他那些排的抱怨。
代际两难并不涉及到同代人。这使之更加难以化解。在这些两难中，如果全都而不是全不停止给予它自身以某些种类的优先权，这对于所有世代而言将更好，第一代除外。那些不同的世代不能够交流和达成条件性的联合协议。先前的那些世代也不会被来自后来那些世代的威胁所吓倒。这个两难中包括局外人这一点因而是一个更大的难题。在一个代际两难——这无需涉及到人口的增长——之中，现存的世代总是处于某一排在先前的诸排站立起来之后所处的位置。它已经遭受到先前的那些世代的行动之苦。而且这个先前的行动现在不能够被任何政治的或者道德的解决之道所改变。既然如此，对于现存的世代而言如果在这样的一种解决办法中发挥作用则会更糟一些。它不会被不愿意无功受禄这一点所促动，因为它不能够从这一解决之道中受益。它会从自己的行动中有所失去，作为报答却一无所获。但是，它较少可能在一种解决办法中发挥作用。那么相同的推理将适用于下一个和所有后继的世代。

[3]参见，例如，Samuelson，第551页。

[4]参见，例如，Hare（1）和（2）中对于道德推理的不同层次的讨论。

[5]Nozick（2），第41页。

[6]Plato（2），21c-d。

[7]按照趋小的边际效用律的某些版本，这就是其中所蕴含的。按照这些版本，如果给予那些境况更糟一些的人们，那么每一资源单位产生更大的效用，所以最有产出的分配将是每人的生活勉强值得过之处的分配。这里有一个明显的失察之处。即便使每个人的生活达到生活开始值得过的那个水平也需要大量的资源。当它们只是用来防止特大量的人们具有那些值得终结的生活（或者具有净负效用）的时候，这样的一些资源并不有助于产生因果上可能的最大的效用净额。

第十八章

荒谬的结论

我们需要一个既解决那个非同一性问题又避免那个令人讨厌的结论的理论。正如我们将要看到的，数个理论能够达到这些目标中的一个，但是不能达成另一个目标。

132.一种据称的非对称性

在许多人看来，我们还有另一个应当达到的目标。考虑一下

> 悲惨的孩子。某个妇女知道，如果她生一个孩子的话，他将罹患多种疾病，以至于他的生活将会糟得无以复加。他将永远不会有什么发展，将会很短命，将遭受无法治愈的疼痛。

即便我们拒斥"糟得无以复加"这个短语，知情的情况下怀上这样一个孩子显然也是错误的。错误之处主要不是来自对他人的影响，而是主要来自这个孩子那种可预测到的骇人听闻的生活质量。

下面回忆一下那个幸福的孩子的事例。我所说的那对夫妻能够假定，如果他们生这个孩子，这对他们或者其他人而言都不会更糟。这两个事例的关系是什么?

一些作者主张，尽管我们有不要生这个悲惨的孩子的责任，但是我所说的那对夫妻却没有什么生那个幸福孩子的责任。他们生育这个孩子的话，只是道德上更好一些罢了。

许多人甚至会否定最后这个主张。这些人认为，尽管生育那个悲惨的孩子会是错误的，但是我所说的那对夫妻却没有什么道德理由生那个幸福的孩子。[1]这个观点被称作非对称性。[2]

如果我们接受这个观点，我们就有第三个目标。除了解决那个非同一性问题和避免那个令人讨厌的结论之外，我们还必须解释那个非对称性。什么理论会实现这些目标?

133.理想契约方法何以提供不出解决方案

我们能够诉诸我将称之为理想契约主义的道德推理方法吗？这得到许多学者的辩护，其中最著名的是罗尔斯。照此观点，最好的道德原则是我们作为社会所遵循的原则合理地择取的那些原则。为了确保不偏不倚，我们要问，如果我们并不知道有关我们自己的那些特定事实的话，我们应当选择什么原则。

一些人主张，如果我们把这个方法应用于那个应当有多少人的问题，我们能够解释我们为什么既应当拒斥那个非人称的总量原则又应当拒斥那个令人讨厌的结论。我们能够解释为什么B会比A更糟，以及为什么在所有的结果中Z是最糟的。即使我们对有关自己的那些特定事实并不知情，我们也会选择一个产生A而非产生B的原则。我们于是肯定境况会更好一些。

正如一些学者所论证的，[3]这并不是用以辩护有关多少数目的人应当生存的原则的一个可以接受的途径。这一道德推理诉诸我们从自利的观点来看所合理地选择的东西。对这个方法而言，我们并不知道是否会承受某个被选原则的冲击这一点是至关重要的。于是，如果所选定的原则会令妇女陷入不利地位的话，我们必须设想我们自己并不知道我们的性别。只有这层无知之幕才使我们的选择像道德所要求的那样不偏不倚。

按照上述推理，选择一个会产生A而非产生B的原则是合理的，因为我们肯定境况将会更好一些。这个推理假定，无论所遵循的是什么原则，我们将肯定存在。这个假定违反了不偏不倚的要求。我们所选择的原则影响到究竟有多少人存在。如果我们假定无论所选择的原则是什么我们都肯定存在，这就如同假定，当选择一个会令妇女陷入不利地位的原则的时候，我们将肯定是男人。

一些学者提出，我们应当改变我们的假设。当我们问什么原则对我们来说是合理的选择的时候，我们应当设想我们并不知道我们将究竟是否存在。如果我们选择某个将会导致一个较小人口的政策的时

候，这将加大我们根本就不会存在的几率。一直有这样的主张，认为随着我们的假定的变更，选择非人称的总量原则会是合理的。

我们不能以这种所提出的方式改变我们的假设。我们能够想象一个不同的、可能的历史，而我们在其中从未存在过。但是我们不能够假定，在事实上的世界历史上，我们根本不存在这一点可能是真的。因此，我们不能够问，照此假定来说什么是合乎理性地选择的。

理想契约方法不能够用于道德的某些方面。于是，正如罗尔斯所主张的，这个方法不能够令人信服地应用于那个我们应当如何对待动物的问题。[4]我认为，出于刚才给出的一个不同理由，这个方法不能够令人信服地应用于有关应当存在多少人的原则的选择。当我们在讨论这个问题的时候，除非我们设想出某种我们不可能设想出的东西，这个方法就不是不偏不倚的。

这个方法的辩护者可能拒斥我的有关不偏不倚的主张。他可能继续认为，在应用这个方法方面，我们能够假定我们肯定将存在。考虑一下人类历史上最后一代人的两种可能性。（考虑最后一代人是为了简化事例。）在太阳爆炸突然终止人类历史之前，历史可能走的路有两条：

> 在地狱甲中，最后一代人由10个无辜的人构成，其中每一个人遭受50年的极痛。这些人的生活糟得无以复加。如果可能的话，他们都会自尽。
>
> 在地狱乙中，最后一代人不是由10个人，而是由10组100万个无辜的人构成，其中每一个人遭受50年缺1天的极痛。

如果我们假定在这两个地狱中的一个我们将肯定存在，从自利的角度看，倾向选择地狱乙显然是合理的，因为我们从而少受一天的极痛。我们应当得出结论说，在道德相关的意思上地狱乙更好一些吗？如果这个地狱就是撒旦所营造的地狱，撒旦在其中的糟糕作为会轻一点吗？

这两个问题的答案都是否定的。地狱乙在一个方面更好一些。每一个人的极痛量会略微少一点，减少量少于0.01%。但是这个事实在道德权重上被那些在极痛中忍受糟得无以复加的生活的人的数目的巨大增长而压倒。在地狱乙中，受苦的量几乎大100万倍。

该例子表明，按照所提出的那个道德推理方法，我们被引导到完全忽略了必须承认至少具有某种道德重要性的东西。我们被引导到忽视了这样一个事实，即在这些结果的一个结果中，会有更为多得多的痛苦。这样的一种道德推理方法是不能够接受的。我们必须另辟蹊径以达到我们的目标。

134.狭义的对人产生影响的原则

如果我们打算避免那个令人讨厌的结论，就必须拒斥非人称的总量原则。我们必须主张，这个原则错误描述了关乎善行或者人的福祉的那部分道德。

我们可能主张说，这个原则是以错误的形式呈现的，因为它仅仅把人当成了价值的储存者或生产者。下面是一段引文，其中这一特征特别明显。在麦凯的著作《幸福的技术》中，他写道：

> 正如需要锅炉利用煤炭的潜能产生蒸汽一样，同样需要有知觉的存在物把某个给定地域中存在的幸福潜质转化为现实的幸福。而且就像那位工程师将选择那些在把蒸汽转化为能量方面最有效率的锅炉一样，正义同样将选择那些在把资源转化为幸福方面最有效率的有知觉的存在物。[5]

这个蒸汽产生模型会被说成是对这部分道德的一种怪诞歪曲。我们会诉诸

> 对人产生影响的限制条件：这部分道德，即关乎人的福祉的那部分道德，应当完全以对于那些受到我们行动影响的人们来说是好还是糟的角度来解释。

这是纳维森所提出的观点。[6]照其观点，不是因为人的生活包含幸福，人存在才是好的，而是因为幸福对人是好的，所以幸福才是好的。

纳维森还主张，通过造成某个人存在，我们并不能给这个人带来益处。他因而主张说，就增加幸福总量的两个途径——造成人幸福和造成幸福的人——来说，只有前者对人而言是好的。他主张说，既然这部分道德只关乎对人来说什么是好的或者什么是糟的，第二种增加幸福的途径，在道德上是中性的。

正如我在第十六章中所论证的，我们必须拒斥有关糟糕的东西必定对某个人而言是糟糕的那种观点。可能看上去，在拒斥这个观点时，我们是在拒斥对人产生影响的限制条件。情况并非如此。与纳维森不同，我们可能认为，通过造成某个具有值得过的生活的人存在，我们从而有益于这个人。"附录"G辩护的就是这个信念。如果我们接受这个信念，我们能够从对人产生影响的角度解释，为什么我们具有道德理由而不去产生某些结果，即便这些结果对没有什么人而言是糟糕的。我们能够解释我们为什么具有不去选择冒险政策或者消耗的道德理由。

如果我们认为使某人存在能够带来益处，我们就必须在3个不同的善行原则之间作出选择。与功利原则不同，这3个原则并不主张涵盖整个道德。既然它们包含"如果在其他事情同等的情况下"这个短语，它们容许我们诉诸其他原则，诸如平等原则。我们还可能主张，当给他人带来益处需要我们付出太多的牺牲或者对我们的生活干扰太大的时候，我们没有义务给他人带来益处。

这3个原则都主张：

（1）如果使某人存在，而且具有一种值得过的生活，这个人从而是得到益处的。如果这个人的生活更值得过，这个益处就更大。

（2）如果在其他事情同等的情况下，故意地做出某个会使结果更糟的选择是错误的。

（3）如果在其他事情同等的情况下，倘若两个结果中的一个结果对人们来说更糟，它就是更糟的。

这些原则在有关什么造成一个结果对人们而言更糟的主张方面存在差异。假设我们在比较结果X和Y。我们将在结果X中存在的那些人称为X-人。在这两个结果中，称X

狭义上"对人们而言更糟"，如果X而不是Y发生对X-人来说更糟的话。

根据主张（1），通过使人存在，某个人能够从中得益。正如我在"附录"G中所写到的，我们无需主张这个结果比替代选项对这个人而言更好一些。这会蕴含着这样一个不能令人信服的主张，即如果这个人从未存在过的话，对这个人来说更糟。代之我们能够主张，如果使某个人存在，对这个人而言能够是好的。我们能够同样主张，如果使悲惨的孩子存在，对他来说这是糟糕的。此乃为什么在刚刚给出的定义中我纳入了"或者对……糟糕"这个短语的原因。

如果我们在刚刚定义的意义上使用"对人们而言更糟"，主张（1）到主张（3）陈述了狭义的对人产生影响的善行原则，简称为狭义原则。

135.我们为什么不能诉诸这个原则

狭义原则在直觉上是令人信服的。如果某个选择对活着的任何人而言都不是糟糕的，我们的善行原则就不应当谴责这个选择，作出这样的假定是很自然的。但是，如果我们诉诸那个出于对人产生影响的限制条件，我们必定拒斥狭义原则。这个原则不能够解决那个非同一性问题。我们需要解释为什么我们具有某个不去选择那个冒险政策或者消耗的道德理由。根据狭义原则，我们没有这样的理由，因为我们知道这些选择对没有什么人来说更糟。

对狭义原则还有一些别的诘难。其中之一是，如果这就是我们的善行原则，我们必须部分接受那个令人讨厌的结论。如果所发生的是Z而不是A，这对活着的任何一个人来说既不会更糟，也不会是糟糕

的。如果我们既诉诸那个出于对人产生影响的限制条件，又诉诸狭义原则，我们必须主张Z比A并不会更糟。我们中的大多数人会发现这难以令人相信。

另一个诘难是狭义原则可能蕴含一些矛盾之处。考虑一下这两个结果：

（1）与其他现存的人不同，杰克具有一种糟得无以复加的生活。吉尔根本不存在。

（2）与其他现存的人不同，吉尔具有一种糟得无以复加的生活。杰克根本不存在。

如果所发生的是（1）而不是（2），对杰克来说这将是糟糕的，而且对吉尔来说并非是好的。因而对（1）之人来说（1）比（2）更糟。同样的推理，对（2）之人来说（2）比（1）更糟。狭义原则蕴含着一个矛盾的结论，即这两个结果中的每一个都会比另一个糟糕。

将对人产生影响的限制条件和狭义原则结合起来，这颇具诱惑力。但是，如果我们诉诸这个观点，我们不能够解决非同一性问题，我们必须部分接受令人讨厌的结论，而且我们被导向矛盾。我们因而已经得悉，我们不能够诉诸这个观点。这是进步，却是一种否定的进步。

我们还在另一个方面取得了否定的进步。许多人接受非对称性。这些人想要解释，虽则在知情的情况下生那个悲惨的孩子会是错误的，但为什么我所说的那对夫妻还是没有什么道德理由生那个幸福的孩子？鉴于我们不能诉诸刚刚提到的那个观点，我只在一个尾注中论证这个观点本会提供非对称性的最好解释。[7]

136.两个广义的对人产生影响的原则

现在我再表述两个对人产生影响的善行原则。像狭义原则一样，这两个原则都主张：

（1）如果使某个人存在，而且具有值得过的生活，这个人从而得到了益处。如果这个人的生活更值得过的话，这个益处就更大。

（2）如果在其他事情同等的情况下，在知情的情况下作出将使结果更糟的选择是错误的。

（3）如果在其他事情同等的情况下，如果两个结果中的一个对人们而言更糟，它就是更糟的结果。

称结果X

在广义上“对人们而言更糟”，如果X的发生对X-人而言比Y的发生对Y-人而言少好一些的话。

在不同数目选择中，“对……少好一些”是模棱两可的。称X

在广义总量的意义上“对人们而言更糟”，如果X之发生所带给X-人的净利益总量会少于Y之发生所带给Y-人的净利益总量的话。

称X

在广义均量的意义上“对人们而言更糟”，如果X之发生所带给X-人的人均净利益均量少于Y之发生所带给Y-人的人均净利益均量的话。

如果我们在广义总量的意义上使用“对人们而言更糟”，主张（1）到（3）陈述的是广义总量的对人产生影响的原则。如果我们在广义均量的意义上使用“对人们而言更糟”，（1）到（3）陈述的是广义均量的对人产生影响的原则。（正如我在注释[8]中所解释的，如果我们拒斥主张（1），否认使某人存在能够带来益处，那么这两个广义原则与狭义原则相吻合。）

照广义总量原则，最好的结果是给人们带来最大净利益总量——减去负担之后的最大利益量——的结果。照广义均量原则，最好的结果是给人们带来人均净利益最大均量的结果。通过诉诸这些原则，我

们能够以第25节中所辩护的那一方式扩大"利益"的用法。最有益于人们的行动是其结果带给人们最大净利益总量或者最大净利益均量的行动。至于许多别的行动也将是获得这些利益的原因中的一些部分这一点并不相关。

两个广义原则都解决非同一性问题。两者都蕴含（Q），即相同数目质量主张。而且这些原则以一种更为人们所熟悉的方式解释了这个主张。我们不去选择消耗的道德理由是什么？照这些广义原则，这个选择有益于那些后来活着的人们，因为他们的生活是值得过的，而且他们的存在多亏了我们的选择。但是，如果我们本来选择保护的话，后来活着的就会是其他一些人，而这其他一些人本会享有更高的生活质量。照这些广义原则，如果使人们存在，而且享有更高的生活质量，这些人就受益更多。对消耗的诘难是，它尽管有益于后来活着的人们，但是它有益于这些人们之处要少于保护会有益于那些本会后来活着的人们之处。那个14岁的女孩为什么要等到后来再生孩子？因为现在生孩子所有益于这个孩子之处很可能要少于后来生孩子所有益于另一个孩子之处。类似的主张适用于我们对那项冒险政策的选择。这个选择有益于那些后来在灾难中丧生的人们，但是它有益于这些人之处要少于安全政策有益于本会以后生活的另一些人之处。

这些主张并不涉及我早先所拒斥的那类文字把戏。当我们讨论那个14岁的女孩的时候，"她的孩子"和"他"这些语词能够用来涵盖这个女孩本会生的所有不同的孩子。在这点所提供的一个为真的意义上，如果这个女孩现在生她的孩子的话，对他而言将更糟。但是在这个主张为真的那个意义上，它并非关涉对特定的人们而言什么是好的或者什么是糟的。这个主张因而必定诉诸某个新的原则，这有待解释和证明。

两个广义原则并非如此。就像我们刚刚看到的，当我们在两个广义的意义上称两个结果中的一个"对人们而言更糟"的时候，我们的主

张却是关涉对特定人们而言什么会是好的或者什么会是糟的。对广义原则人们并非全然熟悉。不能够把它们说成是我们有关人们利益影响的日常原则——我所称的我们的善行原则。这是因为我们现在是在假定，通过使某人存在我们能够从而有益于这个人。我们的日常原则并不告诉我们这样的一些益处的道德重要性何在。我们需要扩大我们的日常原则，以便它回答这个问题。而且至少有3个回答值得考虑。这些就是两个广义原则和那个狭义原则所提供的回答。因而我们能够主张说，这些并非某种有待解释和证明的新原则。它们陈述了我们能够令人信服地展开我们的日常原则的3个途径，这是倘若我们假定使某人存在能够带来益处所需要做的。

既然两个广义原则是我们的日常原则的一些扩展，它们解决了相同数目选择中的非同一性问题。它们以一种令人满意的方式解释了主张Q。但是这并不表明，这些原则中的任何一个是可以接受的。我们必须问，它们在不同数目选择中蕴含着什么。

回到A和B的不同结果。与A相比，B中有两倍的人，他们全都境况更糟。但是B-人的生活值得过的程度相当于A-人的生活值得过的程度的一半多。如果出现的是B而不是A，这在广义总量的意义上会“对人们更好”。较之A对于A-人中的每一个人而言，B对于B-人中的每一个人而言会少好一些。但是既然每一个B-人享有A-人所享有的一半多的利益，而B-人的数目有两倍，他们合起来得到的利益更多，或者会得到更大的利益总量。（在同样的意义上，如果我们多赋予两个人各4年寿命，他们合起来所获得的利益要大于我们多赋予了5年寿命的某个第三个人所获得的利益。）广义总量原则于是蕴含着，如果在其他事情同等的情况下，B会比A更好。

同理，C会比B更好。而Z会是最好的。如果Z出现的话，Z-人中的每一个人得到的利益从而会微乎其微。但是，如果Z足够大的话，Z-人合起来会获得最大的利益总量，或者说受益最多。广义总量原则于是

蕴含着那个令人讨厌的结论。既然我们想要避免这个结论，我们必须主张说，应当拒斥这个原则。

正如我谈及它们时所陈述的，这两个广义的对人产生影响的原则扩展了我们对"利益"一词的日常用法。或许遭到诘难说，通过扩展"利益"的使用，我的原则并不完全满足那个对人产生影响的约束条件。我们会诉诸一些更接近我们对"利益"的日常用法的不同的版本。我们会主张

> 如果我们做的是X而不是Y，而倘若我们做X而不是Y带给X-人的净利益总量大于我们做Y而不是X带给Y-人的净利益总量，这在广义总量的意义上更有益于人们；

而且

> 如果我们做的是X而不是Y，而倘若我们做X而不是Y带给X-人的净利益均量大于我们做Y而不是X带给Y-人的净利益均量，这在广义均量的意义上更有益于人们。

我们然后会主张说，如果在其他事情同等的情况下，当造成两个结果中的一个会更有益于人们的时候，这个结果会更好。这两个定义带给我们广义总量对人产生影响的原则和广义均量对人产生影响的原则的不同版本。

这个版本的广义总量原则，也蕴含着那个令人讨厌的结论。如果有足够的Z-人的话，造成Z而不是A的出现所带给Z-人的净利益总量，大于造成A而不是Z的产生所带给A-人的净利益总量。令人惊讶的是，这个版本的广义均量原则也能够蕴含着那个令人讨厌的结论。假设我们导致了A的出现。我们现在或许面临新的选择。假设短期内我们可以把A变成B。如果我们作这样的改变的话，在新的广义均量的意义上"对人们而言更好"。这个改变会使现存人口增加同样多的新的人。原先就存在着的那一半人的生活质量会出现下滑。这个变化对他们来说

会更糟。但是这个变化会带给新存在的那一半人更大的利益。从A变为B因而会带给B-人一个人均净利益均量。而且它所带给他们的人均利益均量或许大于未能造成这个改变的情况下所带给A-人的人均利益均量。

为了具体说明此点，我妄作精确。假定A中的水准值为100点，B中的水准值是76点。在把A变为B的过程中，原先就存在的那一半人中的每一个人都失去24点，而新存在的那一半人中的每一个人都得到76点。从A变为B因而带给B-人一个净利益均量，即（76-24）/2，或者说26点。未能造成这样的变化会带给每一个A-人24点的利益。把A变为B从而带给B-人的人均利益要大于未能造成这样的变化会带给A-人的人均利益。按照上述的定义，广义均量原则蕴含着，B会比A更好。同理，C会比B更好，D比C更好，E比D更好，依次类推。这些和类似的主张间接蕴含着Z会比A更好。

对这些原则的另一个诘难是，他们时常蕴含着矛盾。在我们刚刚讨论的事例中，广义均量原则的确如此。假设，相较于水准值在76点的B的人口，C会有水准值在58点的两倍的人口。正如我业已表明的，照广义均量原则，B会比A更好。依次类推，C会比B更好。这间接地蕴含着，C会比A更好。但是广义均量原则直接蕴含着A比C更好。根据这个原则，A既会比C更好，又会比C更糟。我在注释[9]中证明此点。

我们可以修正这些原则以避免这样的一些矛盾。但是，正如那条注解所解释的，那些原则因之是不适当的。既然的确如此，而且这两个原则都蕴含着那个令人讨厌的结论，因此广义原则的这两个版本劣于早先所陈述的那些版本。在下文中我提及这些早先的版本。

我们再次取得进步，一种否定的进步。令人讨厌的结论蕴含在非人称的总量原则中。在其快乐主义的版本中，这个原则主张，如果在其他事情同等的情况下，最好的结果是幸福净量最大的那个结果。这个原则的另一个版本用“任何使生活值得过的东西”替代“幸福”。

既然我们想要避免那个令人讨厌的结论，我们必须主张应当拒斥非人称的总量原则。纳维森提出了一个拒斥这个原则的根据。我们能够诉诸对人产生影响的限制条件：任何善行原则都必定呈现出一种对人产生影响的形式的主张。

我们现在已经得悉，诉诸这个限制条件不果，而且更难以解释为什么我们应当拒斥总量原则。首先假设，如果我们造成某个将享有值得过的生活的人的存在，我们并非从而有益于这个人。如果这样的话，我们不能够诉诸对人产生影响的限制条件，因为那么我们就不能解决非同一性问题。照这个替代选择，我们不能够以下面这样的主张拒斥非人称的总量原则：既然它不是对人产生影响的，它所呈现出的形式——牛奶生产模型——是错误的。为了解决非同一性问题，我们必须诉诸一个并非对人产生影响的原则。

其次假设，通过使某人存在，我们能够从而有益于这个人。如果是这样，我们能够诉诸对人产生影响的限制条件。但是这一无所获。广义总量原则以一种对人产生影响的形式重述了非人称原则。为了避免那个令人讨厌的结论，我们必须主张，我们应当拒斥这个原则。而且更加难以解释：为什么应当这么做。既然这个原则是对人产生影响的，我们不能够主张说它呈现错误的形式：把人们仅仅当作价值的储存者或者生产者。更加容易否认：如果幸福更多或者任何造成生活值得过的东西更多，就会更好一些。更加难以否认：如果人们得益更多的话会更好一些。

尽管更加难以否认广义总量原则，但是我们并非被迫接受这个原则。否认这个原则的一些主张中的这样一个主张是可以得到辩护的：即有关使某人存在能够带来益处的主张。而且，即便我们认为存在能够带来益处，至少还有另一个对人产生影响的原则：广义均量原则。这个原则并不蕴含那个令人讨厌的结论。如果所出现的是Z而不是A，这在广义均量的意义上“对人们而言更糟”。Z-人较之于A-人，所获得的利益人均会较小一些。

广义均量原则从对人产生影响的角度重述非人称的均量原则。我们需要一个新的善行原则。我们需要这样一个原则，它既解决非同一性问题，又避免那个令人讨厌的结论。我们可能还想要这个原则解释那个非对称性。均量原则在其非人称形式和对人产生影响的形式这两者中都实现了前两个目标，而且部分地达成了第三个目标。[10]这个原则是我们所想要的对善行给予最佳说明的原则吗？

137.一些可能的理论

均量原则只是取得我们那些目标的原则中的一个原则。当我表述完这些选择余地中的某些替代选择的时候，判断这个原则将更加容易。

我们不能以这个原则呈现出失当的形式这一主张来拒斥广义总量原则。但是我们能够主张，这个原则对我上面所提的一个问题的回答是错误的。我问的是，“在某个时期，生活的质与幸福的量及其他任何一种使生活值得过的东西的量的相对价值是什么？”在可能的那些回答的变域中，总量原则处于变域的一个极端。照这个原则，无论照其非人称的形式还是照其对人产生影响的形式，只有量具有价值。

按照一个较不那么极端的观点，质和量两者都具有价值。照此观点，如果幸福量相同，如果这种幸福呈现在更多的而质更低的生活之中，那么，这种幸福将更糟。要想使B不比A更糟，它不得不包含的可能不是至少相同的幸福总量，而是一个更大的总量，大概是两倍、十倍甚至更多倍。关于任何一种使生活值得过的东西的量，可以作出类似的主张。

另一个极端的观点是认为只有质具有价值。照此观点，如果人们处境更糟起来，那么在道德权重上则永远不能被有再多的人活着这样一个事实所超越。质的方面的折损在权重上永远不能被量的方面的增益所超越。均量原则就是这个观点的一种版本。

不仅在关于某个时期不同的人口方面，而且在关于到底应当有多少人存在这个更大的问题方面，都可能有人持这个观点。正如我早已主张的，均量观点应当呈现暂时的中立形式。以这种形式，它回答这个更大的问题。

假设生活过的人的数目非常小。在一个想象性的历史中，亚当和夏娃享有非常值得过的生活，而且没有生儿育女。他们是惟一活着的人。在一个与之不同的可能历史中，数以百万计的人享有非常值得过的生活，但是没有任何一个人享有像亚当和夏娃那样好的生活。在这第二种可能的历史中，生活质量稍微低一点，但是幸福及别的任何一种使生活值得过的东西的量要大得多的多。照均量原则，或者照那个只有量重要的更加简单的主张，第一种可能的历史会更好。有一些人接受这个结论。

其他人会接受这个主张的那些较温和的版本。既然我们想要避免那个令人讨厌的结论，区分两组答案将有所助益。就像下面页上所示意的那样，这些答案可以由两个问题引入（见下页）。

（1）是总量原则。均量原则是（7）的一个版本。在此忽略这7个观点中的一些观点之间的差异，我们能够承受得起。我认为（2）比（1）更令人信服。而且我认为（5）比（6）更令人信服，而后者比（7）更令人信服。但是，鉴于我们的目的，我们可以忽略这些细节。[11]

我们可以忽略（1）和（2）之间的差异，因为两者对我的第一个问题的回答都是肯定的。两个观点都主张，既然量的价值没有上限，生活质上的任何损失都会在道德权重上被量上的充分增益所超过。这两个观点从而都蕴含那个令人讨厌的结论。差异只在于，照观点（2），Z的人口必定更大一些。这些观点是那个有关量总会重于质的主张的变体。

（3）到（7）是对这个观点加以否定的那些变体。根据（3）到（7），量并不能总是重于质。

我们可以忽略（5）和（6）之间的差异。我们可以把这些观点当作那个有关量的价值具有上限的观点的变体。（7）是这个上限总是为零的情况下的那个变体。（5）到（7）只有当世界人口小的时候才会出现不一致。那么（5）和（6）主张，更大的量会有某种价值，而（7）则否定这种观点。

如果有一个较低的生活质量，它总是在道德权重上比那些存在着并享有值得过的生活的人的数目的增益所超过吗？更确切地说，某个时期质的折损总是被幸福及其他任何一种使生活值得过的东西的量超过吗？

是

否

在任何一个时期，质和量的相对价值是什么？

只有量具有价值。（1）

两者总是具有价值。（2）

量在那些其质低于一定水平的生活中没有什么价值。（3）

在那些其质低于一定水平的生活中量的价值没有限度，但是没有这个价值的什么数量会像在那些其质高于一定水平的生活中量的价值的某个数量一样好。（4）

只有质具有价值。（7）

量的价值有个上限，而且在当今的世界这个限度已经达到。（6）

量的价值永远不能够超过一定的上限。随着量的实际水平的增加，超额的量的价值下滑并且渐趋于零。当量的水平更高的时候，量的价值的上限就更高。（5）

照观点（5），在质的任何水准上，超额的量的价值渐趋于零。这意味着，量的任何超额单位的价值越来越接近零，但事实上永远不

会达到零。这个主张并不蕴含着，量的价值具有一些上限。假设每一个值得过的超额生活的价值如下：1，1/2，1/3，1/4，1/5，1/6等等。每一个超额生活的价值此处接近零；但是这样的一个数列的值没有限度。既然（5）主张有这样的限度，每一个超额生活的价值必定以一种与之不同的方式递减。例子之一会是：1，1/2，1/4，1/8，1/16等等。尽管每一个超额的生活总有某种价值，这个数列有一个上限：2。但是，因超额生活的有穷数目，这个上限根本达不到。

现在世界的实际人口达到几十亿，其中的绝大多数认为他们的生活是值得过的。那些接受观点（5）的人中的绝大多数人会同意，在目前的世界上超额量的价值可以被当作已经达到其限度。这就是我之所以假设在我想象出的A结果中会有100亿人口生活的原因。这使我的例子与实际上的世界相关地相似；使其成为观点（5）和（7）都认同的一个例子。

在下文中我还将忽略（5）和（6）之间的差异。（5）更令人信服。但是，当实际人口非常大的时候，这两个观点之间的差异没有什么实际意义。如果量的价值极端接近其上限，它能够被当作已经达到其上限。通过主张存在一个单一的上限，（6）再次比（5）更不令人信服。但是这个差异在下文中并不相关。我们因而能够承受得起只讨论更简单一些的（6）。

观点（6）能够通过下面这个观点得到表达

有限量的观点：如果在任何一个时期，幸福的净总量小于本会有的量，或者任何一种使生活值得过的东西的总量更小一些，那么，除非这个总量高于一定的限度，否则就更糟糕。

下面假设我们拒斥观点（1），因为我们认为质具有价值。我们认为，如果生活质量低一些的话，总会是糟糕的。用快乐主义的话来说，如果同一个幸福总量更加稀少地分布，或者在更多的生活之间予以分配，结果每一个生活中的幸福更少的话，会是糟糕的。在其最简单的形式上，这个信念能够通过下面这个观点得到表达

两水准质的观点：如果那些活着的人，与本会存在的所有那些人相比，都将更糟，或者具有更低的生活质量，那就将更糟糕。

照此观点，如果在两个结果的一个结果中每一个人都会更糟的话，那么，这就是这个结果的一个糟糕的特征。但是据此推论不出在这两个结果中，这个结果更糟。这个糟的特征可能被一个好的特征所超过。由于两水准质的观点只是有关一个特征之糟糕性的主张，我认为我们绝大多数人会接受这个观点。

假设我们既接受这个观点又接受有限量的观点。在涉及小的人口的那些情况下，我们将认为低一些的质会被更大一些的量所超过。但是，如果量的价值被当作已经达到其限度，那么，我们认为只有质具有价值。正如我所说过的，这正是我们绝大多数人有关20世纪的现实世界所认为的。而且这会是我们关于我所想象出的结果A和B所认为的。由于我们会认为，在A中，量的价值能够被当作已经达到其上限，所以我们会认为B会比A更糟。无论B中包含的人多多少，也会是如此。当考虑多达上百亿人口的时候，我们将认为，质的折损根本不会在道德权重上被量上的增益所超过。如果不是100亿人的一组人，而是更多的人都会更糟的话，那么总是会更糟。即使在两个结果中生活质量都非常高的情况下，这也是更糟。（如果我们拒斥这最后一个主张，我们应当转向观点（3），这个观点在第139节中讨论。）

两水准质的观点不会涵盖许多实际的事例。在由两个人口增长速率所产生的两个结果里的一个结果中，每一个人都比另一个结果中的每一个人处境糟糕，此会为真的情况很少见。如果我们认为质具有价值，我们必须提出某个别的主张，用以涵盖大多数实际情况。均量原则就是一个这样的主张，无论其对人产生影响的形式还是其非人称的形式都是如此。许多人诉诸这个观点。但是，正如我将要论证的，我们应当拒斥均量观点。（那些诉诸这个观点的人们可能接受了两水准质的观点，然后匆忙地把“都”换成“平均说来”。）

为了涵盖大多数实际情况，我们会需要某个更加复杂的主张。但是我们可以忽略这些复杂之处。单考虑我刚刚表述的那两个观点就将足够了。其中的一个观点是有关量的价值在任何一个时期都有一个上限的观点。另一个是有关质具有价值的观点。

我在寻求X理论，这是一个既解决那个非同一性问题又避免那个令人讨厌的结论的有关善行的新理论。更一般地说，X理论会是有关善行的最好理论。当应用于我们所作出的所有的选择的时候，包括应用于那些既影响到未来人的同一性又影响到他们的数目的选择的时候，它会有一些可以接受的蕴含。在一个完全的道德理论中，我们不能够回避那个到底应当有多少人活着的问题。但是我暂时搁置了这个问题。我在讨论有关某些时期中的不同结果我们应当相信什么。

我们可能希望X理论会是简单的。如果我们应该接受那个令人讨厌的结论，那么这个理论会是简单的：它会是只有量具有价值这样一种主张。这个主张得到总量原则的所有版本的表达。但是我认为，我们应当拒斥那个令人讨厌的结论。如果是这样，我们会诉诸另一个简单理论：只有质具有价值的主张。均量原则就是这个主张的一个版本。但是我认为，就像其反面一样，这个主张过于极端。我认为，有关善行的最佳理论必定主张质和量两者都有价值。

如果这些信念是得到正当性证明的，那么，最佳理论或许就是我刚刚表述的那两个理论的一种结合。照这个理论，质总是具有价值，而且量在任何一个时期都不能够高于某个上限的意义上具有价值。这个理论在我的那些简化了的例子中达到了我一直力图达到的那两个结果。它解决了那个非同一性问题，而且避免了那个令人讨厌的结论。如果这个理论是可以得到辩护的，它或许是X的一个简化版本：有关善行的最佳理论。

这个理论是可以得到辩护的吗？以它目前的形式不行。

138.受苦总量

再考虑一下

> 两个地狱。在地狱甲中，最后一代人由10个无辜的人构成，其中每一个人遭受50年的极痛。这些人的生活糟得无以复加。如果可能的话，他们都会自尽。在地狱乙中，最后一代人不是由10个人，而是由10组100万个无辜的人构成，其中每一个人遭受50年缺1天的极痛。

当我们考虑这些想象出的地狱的时候，我们不能够令人信服地只诉诸两水准质的观点，或者均量原则。照这些观点，地狱甲要更糟，因为那10个人的生活都比那10个100万人的生活要更糟一些。我们可以同意，在这个方面，地狱甲会更糟。这个地狱中的每一个人不得不忍受略微多一点的受苦，这更糟一些。但是也可以主张说，在另一方面地狱乙会更糟一些。在这个地狱中，受苦总量几乎大100万倍。而且可以主张说，受苦总量的这种巨大增加在道德上超过了每个人生活中受苦总量的微小减少。

我们能够否定这些主张吗？我们得诉诸

> 有限受苦原则：如果在任何一个时刻，受苦总量大于本来会有的总量，除非这个更大的总量高于一定的限度，否则就将是糟糕的。

照此观点，随着受苦总量的增加，额外受苦就不再那么重要。而且这种贬值趋于贬到零。超过某个一定的点，出于实际目的而言，额外受苦不再重要。如果更多的人不得不忍受甚至最为极端的极痛，这一点也不糟糕。

这个观点非常不能令人信服。它比那个令人讨厌的结论都更不可取。当我们考虑受苦的糟糕程度的时候，我们应当主张，这个糟糕程度没有什么上限。如果额外多出一个人不得不忍受极度痛苦的话，这总是糟糕的。无论多少别的人有相似的生活，这也总是同样糟糕。额外受苦的糟糕程度永远不会降低。

在受苦的情况下，量的贬值没有什么上限。我们认为当考虑受苦的时候只有量重要吗？我们绝大多数人不这么认为。我们总会认为，同一受苦量如果更为稀薄地分摊在更多数目的生活上面，会更好一些。这就是两水准质的观点在一些简单的情况下所蕴含的。

我们应当把这些主张添加到我所表述的理论上面。我们的理论现在结合了三个观点。我们认为，质总是具有价值：如果人们比本会有的情况更糟一些的话，就总归是糟糕的。我们还认为，幸福的量和无论别的什么使生活值得过的东西的量具有价值，但是这个价值在任何一个时期都有一个上限。而且现在我们认为，在受苦量和无论别的什么使生活值得终结的东西的量这两者之中都存在贬值或者糟糕。我们认为，这种糟糕没有什么限度。

这个新理论蕴含着

> 荒诞不经的结论：如果有100亿人生活，生活质量都相当于世界现有人口的平均生活质量，那么必定有某个可以想象的更大的人口，其存在会更糟，即便其成员都具有一种高得多的生活质量。这个更大的人口会更糟糕，是因为在这些生活的每一个生活中，会有某种剧烈的受苦。[12]

即使在那些质量比我们的高得多的生活中，也会有某种剧烈的受苦。在那个想象出来的更大的人口中，每一个人都具有这样的一种生活。这些生活会相当值得过，因为这个受苦会被幸福和其他使生活值得过的东西大大地超过。但是，照我们的新理论，这个受苦不能够被这些别的东西在道德上超过。鉴于人口的庞大及其生活的高质量，额外的量没有什么正面的价值。每一个额外生活中的受苦都是一个糟糕的特征，造成结果更糟。照我们的理论，这个糟糕的特征不能在道德上被每一个额外生活大得多的幸福或者其他使这些生活相当值得过的东西所超过。这不是些使结果更好的特征，因为量的肯定价值业已达到其极限。

照我们的理论，任何一个额外的生活都只有两个道德上相关的特征。一个会是这个生活的质量，另一个会是它所包含的受苦。那个荒诞不经的结论比较了两个可能的人口。一个包含100亿人口，所有的人都具有与现在实际上生活着的人的平均生活质量大约相当的生活质量。另一个人口要大得多，其所有成员都享有一种高得多的生活质量。（我们可以假设，这个更大人口的绝大部分生活在业已漂移到太阳系之中并成为太阳系组成部分的一些类-地球的星球上。）较之于那个较小的人口生存，这个较大人口的生存会在某个方面更好一些，在另一方面更糟一些。这个结果会有的那个好的特征是，每一个人都享有一种高得多的生活质量；这个结果会有的那个糟特征是，会有一个更大的剧烈受苦的总量。假如我们想象这个人口在生活质量相同的情况下更大一些，那么，这不会使那个好特征更好。但是它会使那个糟的特征更糟，因为会有一个更大的剧烈受苦的总量。照我们的观点，这样的受苦的一个更大的总量的糟糕没有什么限度。既然这个特征的糟糕没有什么限度，那么它以一个足够大的人口必定超过那个好的特征。我们的新理论显然蕴含着那个荒诞不经的结论。

我们如何能够避免这个结论？我们不能够令人信服地给受苦总量的糟糕设置一个上限。但是我们可以区分两种受苦。如果受苦是呈现在值得过的生活之中，那么这个受苦是得到补偿的。如果受苦出现在一个并不值得过的生活中，那么这个受苦是未得补偿的。我们不能够主张说，在这两种受苦之间，只有第二种是糟糕的。假设在一个一直值得过的生活的末期，某个人开始剧烈地受苦。我们不能够可辩护地主张说，当这个人正在剧烈受苦的时候，这并不糟糕。但是我们可能主张说

如果有更多未得补偿的受苦，总是糟糕的。这个糟糕没有什么上限。而且，如果一个额外的人受苦，且具有并不值得过的生活，这总是同等地糟糕。这个受苦的糟糕程度不能够被其他人是幸福的这个事实所减少。当我们考虑并不值得过的一个生活中的受苦的糟糕程度的时候，其他人身上发生什么并不相关。[13]

有两种方式，其中可能有更多的得到补偿的受苦：（1）在现在正在过活的一个值得过的生活中存在更多的受苦。（2）或许有一个额外的人存在，享有值得过的生活，但是包含某种受苦。就这两者而言，只有（1）是糟糕的。此乃我们何以能够拒斥那个荒诞不经的结论的理由。那个更大的人口会有更多剧烈的受苦这一点不会是一个糟糕的特征。这之所以不会是一个糟糕的特征，是因为这个受苦会出现在一些值得过的额外生活中。如果这些额外的人存在的话，不会更糟。我们就这个额外的受苦所主张的，并非它会是糟糕的，而是，如果这些生活不包含这个受苦的话，本会更好一些。

可能遭到这样的诘难："如果这个额外受苦并不糟糕的话，何以假如这些生活不包含这个受苦的话本会更好一些？"我们会回答说："因为这本会使生活质量甚至更高一些。"

接着可能遭到诘难说："如果你否认那个出现在那些值得过的生活中的额外受苦是糟糕的，你必定是不言而喻地假定，这些生活包含超过该受苦的糟糕性的一些好特征。这从根本上削弱了你有关量的肯定价值业已达到其上限的主张。"

我们会回答说："我们的观点区分了人称性的和道德的价值。我们所称的生活价值不是其个人价值——对拥有这个生活的那个人的价值——而是这个生活贡献给该结果的价值。当我们主张某个生活没有什么价值的时候，这意味着这个生活的过活并不使该结果更好。换言之，如果具有这个生活的人根本不存在的话，这本身并不会更糟。如果这本会更糟的话，也只会是因为这对他人产生的影响。鉴于这一区分，我们能够回应你的诘难。这些额外生活中的受苦具有个人贬值，或者说对那些过着这些生活的人们而言是糟糕的。对这些人们而言，被他们幸福的个人价值和其他使他们的生活值得过的东西的价值超过。这种个人价值并非道德价值。这些额外的人的生存并非使该结果更好。同样，因为它们的受苦是得到补偿的，或者说被同一些生活中的个人价值所超过，这个额外受苦的个人贬值并不具有道德贬值。当只是因为有更多值得过的生活才有更多的受苦的时候，这个额外的受苦并不使该结果更糟。"[14]

下面考虑以下某个时期3种可能的人口。第一种是20世纪的最后1/4世纪期间世界的实际人口。第二种是更大一些的人口，差不多所有的人都具有高得多的一种生活质量。如前所述，这个更大人口的绝大部分生活在业已成为太阳系组成部分的其他类-地球的星球上。

在这个较大的人口中，有一些不幸的人，他们受苦而且生活不值得过。这些人或许就像上文所表述的那个悲惨的孩子，他患有疾病，以至于根本不可能发展，只会活不多的年数，身受不能治愈的疼痛的折磨。或者这些人身受某种令人痛苦的终身精神疾患之苦。这些不幸的人按比例来说相当罕见。在这个想象出的人口中，每100亿人口中会有一个这样的人。

许多人认为，通过折磨一个无辜的孩子而带给数以百万计的人以幸福会是错误的。[15]尽管比例相似，我所想象出的事例却有很大的不同。在这个较大人口的每100亿人中有一个人患有某种疾病，疾病让他受苦，并使其生活并不值得过，但这种情况纯粹是出于糟糕的运气。这些不幸的人并不因为没有发展或者出于精神疾病的本质而自杀。我们中的一些人也许会认为，这样一些人为他们自身之故应该被杀掉。但这并非我所想象出的人们所持的观点。他们认为杀掉这些为数不多的不幸者会是错误的，他们尽其所能缓解这些人的痛苦。

下面假设，因为在每100亿人中只有一个这样的人，这样的一些人在这个更大的人口中要比在目前的现实世界中会少一些。即使这个设想的人口比目前世界人口规模大许多倍，也能够令人信服地如此假设。在世界实际人口中，有一些所患疾病使他们受苦并使其生活不值得过的人们。判断这样的人与那些生活值得过的现实的人们的比例会相当困难。但显然，这个比例远高于每100亿中有一个。

与世界实际人口相比，这个更大的人口在两个方面更好，没有什么方面更糟。绝大多数人的生活质量高得多，而且其中未得补偿的受苦更少。与世界实际人口的生存相比，这个设想的人口的生存显然会好得多。

现在考虑第二个设想的人口。除了大得多的多之外，这个人口完全像第一个人口。这个人口会生活在许许多多的类-地球的星球上。与世界实际人口的生存相比，这第二个人口的生存在一个方面更好，在另一个方面更糟。更好的方面是几乎每一个人都享有一个高得多的生活质量。更糟的方面是未得补偿的受苦会更多。

就像以前一样，如果我们设想这个人口更大一些，它在没有什么方面会更好，在一个方面会更糟。生活质量会是相同的，而且量的肯定价值已经到达其限度。如果这个人口更大一些，它的好特征不会更好。但是它的糟特征会更糟。未得补偿的受苦会更多。而且这个特征的糟糕性没有什么限度。

照我们的理论，这个人口的那些好特征具有一种有限的价值。但是糟糕特征的贬值绝没有什么限度。每当我们设想这个人口更多一些的时候，好特征不会变得更好，但是糟的特征却总是变得更糟。如果这个人口足够大的话，其生存要比世界实际人口的生存更糟。而且，如果这个人口足够大的话，其糟糕特征会超过其好的特征。不受限的糟必然能够超过受限的好。如果这个人口足够地大，其生存会是内在地糟的。如果这个时期没有什么人生存，那就会更好。

我们或许说："即使如果在这个时期没有什么人生存本身会更好，但这从整体上来说更糟，因为那样的话就没有什么未来的人了。"

这不是什么答案。考虑两个可能的未来历史。在第一个中，没有什么未来的人。在第二个中，总是有刚刚表述过的那种人口。既然这样的一个人口的生存，照我们的观点，总是内在地糟糕，我们必然得出结论认为，如果没有什么未来的人的话，会更好一些。

鉴于我们的理论，我们被迫接受

> 荒谬的结论：在一个可能的结果中，在某个未来的世纪既会存在某个就像地球目前的实际人口的地球人口，又存在一个数目庞大的别的人口，他们生活在早已成为太阳系组成部分的那些类-地球的星球上。这些别的星球上的人几乎都会享有远远高于地球实际人口中的绝大多数人所享有的生活质量的一种生活质量。这些别的人中每100亿人总会有一个不幸的人，他患有一种让他受苦的疾病，而且有着一种并不值得过的生活。

在第二个可能的结果中，会有数目与之相同的庞大未来人口，除了每100亿人中的那个不幸的人之外，都具有与之相同的生活质量。但是这个数目庞大的未来人口不会都生活在一个未来世纪中。这些人中每100亿人会生活在许多未来世纪中的一个世纪中。

照我们的观点，第一个结果会非常糟糕，要比倘若没有这些特大量的未来人的情况下要糟糕得多。第二个结果会非常好。即便在两个结果中都有数目相同的特大量的未来人，两者中除了每100亿人中的那个不幸的人之外，所有人都享有相同高质量的生活质量，第一个是非常糟的，第二个是非常好的。

这个结论可能并非荒诞不经。但是其荒谬之处并没有因此而减低多少。除了所有的特大量的未来人都生活在同一个世纪而不是一些不同的世纪之外，第一个结果恰如第二个结果。如果第二个结果会非常好，而第一个结果只在人们什么时候生活方面与之不同，那么，这种时间选择上的差异何以造成第一个结果非常糟？

存在着一种方式，其中时间选择上的差异或许造成道德上的差异。一代人之内的不平等可能比不同代人之间的不平等更糟一些。但是，在我们正在考虑的两个结果中，不平等会是一样的。的确，在两者之中，在每100亿人中会有一个人比其他所有人的处境更糟的。

我们能够假设，在这两个结果中，所过的所有生活在数量上和在质量上两者都是同一的。在这些方面，这些结果确切地相似。但是，因为所过的这些生活在时间上的差异，我们得出结论说，第一个结果会非常好，而第二个结果会非常糟。

那个荒谬的结论显然为我们最新近的观点所蕴含。照这个观点，量的肯定价值在任何一个时期都具有一个上限，而量的负面价值或者未得补偿的受苦数量的糟糕性却没有这样的限度。在第二个结果中，每100亿特大量的人生活在不同的世纪之中，量的肯定价值在每一个世纪都大大地超过其负面价值。这就是这个结果非常好的原因。在第一个结果中，这些特大量的人都生活在同一个世纪中，情况正好相反。既然量的肯定价值在这个世纪有一个上限，这个价值就被未得补偿的受苦数量的糟糕性大大超过。这个结果因而非常糟。即便在两者

中所过的生活在数量上和在质量上都是同一的，第二个结果也非常好，而第一个也非常糟。

我不能够相信这个结论。这个结论得自我们关于量的价值的那些主张中的非对称性。照我们的观点，尽管量的贬值在任何一个时期都没有限度，但是其价值却有一个限度。为了避免这个荒谬的结论，我们必须抛弃这个观点。

139.诉诸没有价值的水准

我们绝大多数人都想避免那个令人讨厌的结论。方式之一是主张量的肯定价值有一个限度。但是，正如我们刚刚看到的，我们必须抛弃这个观点。在第137节的那个观点图示中，我提到两个选择余地：观点（3）和（4）。

观点（3）是

> 诉诸没有价值的水准：量在质低于一定水准的那些生活中没有价值。如果这些生活是值得过的，它们具有个人价值——对那些具有这些生活的人们而言具有价值。但是有这样的一些生活得以过这个事实并不使结果更好。

如果添加下述主张，我们能够使这个观点更加令人信服。没有价值的水准并不涉及一种泾渭分明的断裂。如果生活质量非常高，每一个生活的道德价值——借此它使结果更好——是它的完全的人称性的价值。在某个较低的水准上，一个生活的价值开始比其完全的人称性的价值要少一些，而且在一些较低的水准上这两个价值越来越分道扬镳。在没有价值的水准上，尽管生活仍然值得过活，但是第一个价值已经达到零。说没有什么泾渭分明的间断，是因为在稍微高一点的水准上这个价值接近于零。

我们可以进一步减少这种间断性。我们会主张，在一个值得过的生活中总有价值，但是在某个水准之下，额外量的价值下滑。在这个

水准之下，量的总价值具有一个上限，这个上限可以逼近，但永远不能完全达到。这些主张会使我们的观点更加令人信服。但是它们事实上只会带来微不足道的差异。如果人口足够大的话，低于这个水准的额外量的价值会非常接近于零。既然实际上的差异会是微不足道的，我将讨论并不做出这些进一步主张的那个更加简单的观点。

如果我们诉诸没有价值的水准，那么，我们应当把这个水准当作什么？这个水准是一个宽带的下缘，在这个宽带中，生活的人称性的价值越来越与这个生活贡献给那个结果的价值相分离。如果我们诉诸没有价值的水准，我们需要决定这个宽带出现在生活质量标度的什么地方。我们可能发觉这个问题难以回答。但是这可能比接受那个令人讨厌的结论难度要小一些。既然我们已经被迫放弃有关量的价值具有一个上限的主张，诉诸没有价值的水准可能是避免那个令人讨厌的结论的一个最不难以置信的方式。

除了诉诸没有价值的水准之外，我们应当继续提出我们的其他一些主张。我们应当继续主张，质具有价值：如果生活质量较低总归是糟的。而且我们应当继续主张，如果它们出现在那些并不值得过的生活之内，在受苦和其他别的什么使生活值得终止的东西的量中就有贬值或者糟糕性。未得补偿的受苦的糟糕性没有什么限度。我们现在应当添加这样一个主张，即，倘若这个价值出现在那些高于没有价值的水准之上的生活中，幸福量和其他别的什么使生活值得过的东西的量具有价值。而且我们应当主张，这个价值没有限度。

按照这个新理论，我们避免那个荒谬的结论。在我所表述的第一个结果中，几乎每一个人都具有超出没有价值的水准之上的生活。每100亿人中只有一个人的生活不是值得过的。量的肯定价值在此超过其负面价值。

或许受到诘难说："考虑两个这样的结果，其比例关系与A和B之间的关系相同，但是其中生活质量低许多。与结果J相比，在结果K中会有两倍的人，这些人处境都更糟。下面假设，在这两个结果中，生

活质量都会低于你的没有价值的水准。照你的新理论，在这些结果中所过的生活会没有什么价值，虽然它们会具有人称性的价值，但它们不会具有那种使结果更好的价值。如果在这两种结果中所过的生活不会有什么价值，那么，这蕴含着结果J不会比结果K更好。但是你会主张，既然生活质量在J中会更高一些，J本会比K更好一些。你的新理论是自相矛盾的。”

我们会回答说：“我们主张，如果生活质量低于没有价值的水准，那些所过的生活，即便是值得过的，也没有什么价值。这样的一些生活并不具有量所能够提供的那种价值。我们这么说是指，如果有特大量的人过着这样的生活，这不会使结果更好。仅当生活质量高于没有价值的水准之上的时候，特大的量才有价值。但是特大量的价值并不是一种惟一的价值，或者并非能够使一个结果更好的惟一特征。我们否认只有量具有价值。照我们的理论，质也有价值。我们认为，如果人们处境更好一些或者具有更高一些的生活质量，这总归是好的。既然一个结果能够更好的方式有二，我们的理论就不是自相矛盾的。结果J可能会比结果K更好，不是因为它具有更多的量的价值，而是因为生活质量会更高。”

这个理论提供了逃避那个令人讨厌的结论的一个良好途径吗？当我已经讨论完一个可供选择的余地的时候，这个问题将会更容易回答一些。

140.词典式的观点

代之诉诸没有价值的水准，我们可以诉诸观点（4）。这类似于纽曼有关疼痛与罪的观点。纽曼主张，尽管两者皆糟，但没有多大量的疼会像最小量的罪那样糟。

照这样一种观点，没有单一的价值尺度。尽管疼痛的糟糕没有限度，但是这种没有限度的糟糕不会像另一种有限度的糟糕那样糟。

这样的观点是前后一贯的。当应用于分属两个不同范畴的那些事物的时候，就像在纽曼的事例中的情况一样，它更加令人信服。我们能够把这样一个观点应用于一个单一的范畴——那些值得过的生活吗？

当我们比较人的某些生活和低等动物的生活的时候，持有这样一种观点是令人信服的。如果它们是有意识的，牡蛎也可能具有值得过的生活。当它们不是工场式养殖的时候，猪猡的生活很可能是值得过的。但我们能够令人信服地主张，即便这些生活得以过活这个事实有某种价值，这个价值再大的量也不会像苏格拉底的生活中的价值那样好。

如果我们打算以这种方式避免那个令人讨厌的结论，我们必须做出这样一个关于生活的主张，这些生活都是由人类去过的。照生活质量的尺度，我们必须界定两个新的水准。如果生活低于较低的水准，我们称之为平庸的生活；如果生活高于较高的水准，我们称之为有福的生活。我们现在可以诉诸

词典式的观点：量的实际价值没有限度。如果一个额外的值得过的生活得以过，那总归更好一些。但是平庸的生活的数量再多也不会具有一个有福的生活所具有的那么多的价值。

141.结论

我们必须避免那个荒谬的结论。这个结论得自我们有关量的价值的那些主张中的非对称性。我们设置了某个时期内量的肯定价值的限度，但是没有设置其否定价值的任何限度。为了避免那个荒谬的结论，我们必须放弃这个非对称性。

我们不能够令人信服地设置量的否定价值的一个限度。如果有更多的未得补偿的受苦，那总归是糟糕的，而这个糟糕性永远不会下滑。我们因而必须去除量的否定价值的限度。

当我们去除这个限度的时候，我们需要一种新的方式避免那个令人讨厌的结论。在我的结果A中，会有100亿人生活着，他们都具有像当今活着的最幸运的人们那样高的生活质量。在结果Z中，会有某个庞大的人口，其成员具有勉强值得过的生活。如果这样的生活得以过和量的价值没有限度这个事实有任何一种价值的话，Z会比A更好。 A中量的价值可能非常巨大。但是，如果Z中量的价值没有什么限度，这个价值会更大。这个价值会大到足以超过生活质量低得多这个事实。此外在这个事实中可能存在贬值。但是这个贬值会被一个没有什么限度的价值所超过。如果Z中量的价值没有什么限度，这个价值会超过生活质量低得多这个事实。

这些评述假定，存在一个单一的价值尺度。我们可能会否定这个假定。我们可能诉诸词典式的观点。照此观点，如果一个值得过的特大量生活得以过活，那总归更好一些。我们承认，Z中的量的价值没有什么限度。但是我们主张，这种价值的量再大也不会像一个在有福的水准之上过活的单一生活的价值那样好。我们能够假定，在A中每一个生活都在这个水准之上。照词典式的观点，无论Z的人口规模如何，Z会比A更糟。

我们的可供选择的余地是诉诸没有价值的水准。我们会从而主张，出于实际上的目的，那些在这个水准之下过的生活能够假定为不具有什么价值。如果这些生活值得过的话，那它们就具有个人价值；但是它们并不使结果变得更好。量只有在高于没有价值的水准之上的生活中才有价值。为了避免那个荒谬的结论，我们承认，这个价值没有什么限度。

这些当中哪一个观点更好一些？我们应当采纳词典式的观点，还是转而诉诸没有价值的水准？

如果我们诉诸没有价值的水准，我们不能够避免那个荒谬的结论的变体。我们必须接受

（A）假设在未来的某个历史中，总会有庞大数目的人，而对每一个受苦且生活并不值得过的人而言，会有100亿人的生活是值得过的，尽管他们的生活质量并不怎么高达没有价值的水准。这会比倘若没有什么未来的人的话更糟。

我们必须接受（A），因为照我们的观点，在这个想象出的未来量中不会有什么肯定价值。在为数很少的不幸者的受苦中会有贬值。如果相反没有什么未来的人，这可能可以主张为一个特别的事例，其中会有质的损失。但是这个损失的糟糕程度，无论多么大，都比受苦总量的糟糕程度要小。在有一个庞大的人口的情况下这必定为真，因为这个糟糕程度没有限度。

我们的新理论还蕴含着

（R）如果有100亿人生活着，都具有一个非常高的生活质量，那必定有某个可以想象的更大的人口，其生存会更好一些，即便其成员具有的生活勉强高于那个没有价值的水准之上。

这是那个令人讨厌的结论的一个弱化形式。我们的理论蕴含着这个结论，因为它蕴含着，在那个没有价值的水准之上，任何质的损失总会在道德上被量的一个充分的增益所超过。这必定是蕴含其中的，因为要避免那个荒谬的结论，我们的理论就不再主张量的肯定价值有一个限度。

（A）比那个荒谬的结论的荒谬性要小一些。（R）比那个令人讨厌的结论要较少令人讨厌一些。这两个结论的令人难以置信的程度依赖于我们把那个没有价值的水准设定在何处。我们可能把这个水准定在靠近生活不再值得过的那个水准之处。如果这种没有价值的水准低到这个程度，那么，（A）可能不是荒谬的，但（R）是令人讨厌的。如果我们提高没有价值的水准，那么，（R）不再那么令人讨厌，但是（A）更加荒谬。

如果我们转而诉诸那个词典式的观点，我们必须接受（A）和（R）的变体。这些只是以“平庸的”一词替代了“没有价值的”一词。至于那个词典式的观点何以蕴含着这些结论，我在注释[16]中进行了解释。

在这一章中，我对X理论进行了探寻，这个有关善行的理论解决那个非同一性问题，避免那个令人讨厌的结论，而且具有在所有情况下都可以接受的蕴含。有一个原则达到了前两个目标，这就是广义均量对人产生影响的原则。这个原则主张，如果使某人存在过着值得过的生活，那么他因而得到了益处。正如我在“附录”G中所论证的，这个主张及其否定两者都是可以得到辩护的。

那个广义均量原则本身不可能是X理论。在某些情况下，它具有不可接受的蕴含。于是它蕴含着地狱甲会比地狱乙更糟，即便在地狱乙中每一个生活几乎与之同样地糟，而且未得补偿的受苦数量几乎大100万倍。更宽泛些说，广义均量原则并没有给予这样的受苦总量什么权重。即使我们采纳这个原则，我们也必须加上这样一个主张，即如果未得补偿的受苦更多，而且这个糟糕性没有什么限度，那总归是糟糕的。

当应用于那些值得过的生活的时候，广义均量原则没有赋予幸福量和别的使生活值得过的东西的量什么权重。该原则会蕴含着，在最好的可能历史中，只有一个人算活着。我们绝大多数人会发觉这个观点过于极端。我们绝大多数人会认为，量有某种价值，即使这个价值在任何一个时期都有一个上限。

这个均量原则与有更多限制的两水准质的观点部分契合。根据这个观点，当我们比较两个结果A和B的时候，如果B中的每一个人会比A中的每一个人处境更糟的话，那就是更糟。这个观点涵盖的实际情况没有多少。均量原则以一种使之涵盖所有情况的方式展开了这个观点。但我将很快予以论证，即使作为一种诉诸数个原则的多元理论的

部分，均量原则也是不可接受的。两水准质的观点必须以另一种方式加以展开。如果这些主张是得到正当性证明的，我们就尚未寻获X理论。既然我们应当拒斥均量原则，我们就尚未完全解决那个非同一性问题。

我们还发现了另一个难题。如果我们的理论做出我刚刚表述的那些主张，它以蕴含着那个荒谬的结论的代价避免那个令人讨厌的结论。既然这是一个新难题，我们又取得了一种否定的进步。通过诉诸那个没有价值的水准，或者那个词典式的观点，我们部分化解了这个新难题。我们只是被迫采纳（A）和（R）这两者，或者这些结论的两个变体。这些结论较不那么令人讨厌，也较不那么荒谬。但它们两者都是令人难以置信的。我们并未全部化解这个难题。

正如我现在将要论证的，还有另一个问题。

---

[1]参见Narveson（1）和（2），以及Sikora and Barry中的J·贝内特和M·沃伦。还有许多别的例子。

[2]载McMahan（1）。在整个第四编中，我大大得益于J.McMahan未经发表的著作，以及对这些问题的很多讨论。

[3]参见，例如，Barry（3）。

[4]Rawls，第17页和第504—512页。

[5]Mackaye。我没有具体的参照页数。而且鉴于第二句全凭记忆引述，用词可能并不精确。这本被遗忘的著作会博得功利主义一笑，不仅令他们受益，而且也令他们的对手中意。

[6]载Narveson（1）。近期的辩论之所以存在很大程度上归功于这本开拓性的著作。另参见Narveson（2）。

[7]按照所以为的非对称性，生养那个悲惨的孩子会是错的，但是也没有什么道德理由生养那个幸福的孩子。这如何能够得到解释？

一些学者主张，使某人存在这一点对于这个人而言既不能够是好的也不能够是坏的。如果是这样，这会解释为什么说没有道德理由生那个幸福的孩子。但是它会蕴含着，在生那个悲惨的孩子时，我们不可能在做某种对于他而言糟糕的事。诘难是什么？如果生养这个孩子对于他而言并不是糟糕的，那么它的错误性必定来自这个孩子所受之苦。我们必须诉诸

受苦总量原则：在其他事情同等的情况下，我们不应该增加受苦的总量。

人们可能难以接受这个原则，但会完全拒斥幸福总量原则。但是如果我们接受那个原则，那么我们的那对夫妻就有一个道德理由来生那个幸福的孩子。生这个孩子本身纵然对他而言不能够是好的，但是它会增加幸福总额。

可能有诘难说，受苦和幸福在道德上是不相似的——我们防止前者的道德理由远远重于我们

促进后者的理由。但是这并不能够解释那个非对称性，除非我们没有什么道德理由来促进幸福。而且，如果我们已经接受了受苦总量原则，那么完全拒斥它对幸福的类比看起来就不太可取。（参见Griffin（4））

有一个解释该非对称性的更好方式。我们可以（1）诉诸出于对人产生影响的限制条件，（2）主张使某人存在对于他而言能够是好的或者糟的，并且（3）诉诸那个狭义原则。根据那个狭义原则，倘若其他事情等同，去做那对于存活的人而言是糟糕的或者更糟的事则是错的。生那个悲惨的孩子从而是错的，因为对他而言这会是糟糕的。但是未能生那个幸福的孩子则绝没有什么错。的确，如果我们的那对夫妻生这个孩子，这对他而言将是好的。但是，如果他们并不生这个孩子，这对于他而言并不是糟糕的。而且，在所描述的那个事例中，它对于别的任何人而言也不糟糕。这就是为什么没有什么道德理由来生这个孩子的缘由。

在我看来，这是对该非对称性的最好解释。我们应当注意到，狭义原则并不涉及到人们熟悉的那个主张：我们不去伤害人的义务比我们有益于人的义务更强。我们可能宁愿把这个主张加给我们的狭义善行原则——就像罗尔斯所做的，添加有关非邪恶行径的某个更强的原则。但是狭义原则并不作这样的区分。如果将有某个我们未能给予某种利益的人，那么我们未能做到这一点对于这个人而言将更糟。在倘若其他事情同等的情况下，那么根据狭义原则我们已经错误地行动。

广义原则和狭义原则之间的区分是一个新区分，是由这样一个信念所开启的，就是说，在使某人存在的时候，我们能够做某种对于这个人而言是好的或者糟的事。这个信念打破了这样一个惯常的限定，就是说，假如某个事件对于人们而言是好的，那么这个事件的非发生对人们而言就会更糟。随着这个限定打破，广义原则和狭义原则分道扬镳。

最后我们应当注意到，我们能够证明我们诉诸狭义原则是正当的。按照这个原则，如果在使某人存在方面我们所做的对于这个人而言是糟糕的，那么它则具有道德的重要性；但是，如果在使某人存在方面我们所做的对于这个人而言是好的，那么它则没有什么重要性。J. Sterba已经提出诘难说，狭义原则在此并未解释这个非对称性，因为它只是重述了这个非对称性。这个诘难能够予以回答。我们可以通过诉诸数个有关道德之性质或者道德推理的理论来证明诉诸狭义原则的正当性。Scanlon的理论就是一个例子。正如我所说过的，这理论主张道德动机的来源是"立足他人不能有理地予以拒斥的根据而能够证明一个人给他人所施加的行动是正当的这一愿望”（Scanlon（3），第116页）。按照这个理论，只有当一个行动以这样一种方式影响到某人，而这种方式带给这个人的是一个无法回答的抱怨的时候，这个行动才是错的。我提及可以证明诉诸狭义原则为正当的数个别的理论。按照这些理论，下面这一点在道德上是重要的：如果我所说的那对夫妻未能生那个幸福的孩子，那么将没有什么抱怨者。

尽管它提供了对于这个非对称性的最好解释，但是我们已经看到，我们必须拒斥这个把狭义原则和出于对人产生影响的限制条件结合起来的观点。我们必须拒斥这样一个观点，就是，一个行动面临道德诘难，则必定有某个抱怨者。如果我们选择消耗，这将在后来造成生活质量方面的一个大滑坡。但是那些在这个滑坡之后生活的人们将把其存在归功于我们的选择。既然他们并不后悔他们的存在，他们将不会对我们的选择感到痛惜。将没有什么抱怨者。但对于我们的选择是有一个道德诘难。

[8]如果使某人存在对于这个人而言不能够是好的或者糟的，那么当我们比较两个结果的时候，其利益重要的人惟有那些在两者中都存在的人。所有在一对一对的结果之间的选择，出于道德目的，都像那些相同数目选择，其中狭义原则与广义原则相契合。

[9]在从B到C的变化中，先前存在的那一半各损失18点，而新存在的那一半则各获益58点。因而从B到C的转变给予C-人的净利益均量是每人（58 - 18）÷2，即20。未能作出这个改变会给予B-人每人18点的利益。因而，从B变到C给予C-人带来的人均利益大于未能作出这个改变会给B-人所带来的人均利益。按照上述广义均量原则，C会比B更好。（在对B变到C的平均利益的计算中，我用2相除，因为两群人同等大。这个简便方法与这个原则所明白要求的计算产生相同的结论。本来会要求先把净利益总额减去这数以10亿计的所有人的损失，然后再把这个差除以总人数。同样的说法适用于我的其他简便计算）。既然这个原则蕴含着C会比B更好，而后者会比A更好，那么它间接地蕴含着C会比A更好。但是在从A到C的转变中，先前存在的1/4会损失42点，而新存在的3/4会获益58点。从A到C的变化因而带给C-人的平均净利益是：（58×3 - 42）÷4，即33。从A变到C所带给C-人的人均利益因而小于未能作出这个改变会带给A-人的人均利益。根据这个原则，尽管C间接地比A更好，但是C比A更糟。我们可以借助这样的主张来修正这个原则，就是说，当相关计算产生这样的相互冲突的结果的时候，其中没有哪一个结果会比另一个更糟。这个修正的原则会蕴含着，某些结果会比另一些更糟，但是只会产生一个非常局部的调整，无法支持我们在许多情形中想要的那些答案。（在此处和别处我大大受益于同M·伍德福德之间的通信往来。）

[10]参见McMahan在McMahan（1）中对这个原则的讨论。

[11]观点（5）是由J.McMahan向我提出的。而我们两人又都受到Hurka（3）的影响。观点（1）在下面这个方面与其他所有的观点有所不同。根据观点（1），一个结果的价值是它对于这个结果中的所有人的价值的总和。如果某人具有一个值得过的生活，这个生活对于这个人而言具有价值；而且，如果这个生活更值得过，它则更加具有价值。我称这种价值是人称性的（personal）。按照观点（1），最好的结果就是那个具有最大个人价值的结果。
按照其他6个观点，这是遭到否定的。两个结果之一尽管具有更大的个人价值，但仍然可能更糟。这可能看起来不令人信服。如果一个结果对于这个结果中的那些人具有更大的价值，这看起来可能是据以判定它是那个更好结果的一个强理由。按照观点（2）至（7），一个结果的价值并不对应于其个人价值。在这个意义上，这些观点是非人称性的（impersonal）。这可能看起来是对于这些观点的一个诘难。
尽管在这个意义上只有（1）是一个人称性的观点，但是在另一个意义上（1）则是那个惟一的非人称性的观点。（1）是惟一的并不关心人们的生活质量或者并不关心这些生活中的幸福量的观点。根据（1），如果总量相同，即便人们的生活质量低得多，也并不产生什么道德差异。所有其他的观点都关心人们的生活质量。而且这就是为什么这些观点在第一个意义上是非人称性的——为什么它们否定一个结果的价值对应于其人称性价值总额这个主张的缘由。既然它们因为关心人们的生活质量而否定这个主张，那么在第一个意义上它们是非人称性的这一点就不是什么对这些观点的诘难。

[12]我展开了R.Sikora所给出的一个论证，载Sikora and Barry。

[13]在此像非功利主义者所做的那样诉诸个人的分离性是令人信服的。在其他生命中所发生的东西绝不能补偿或者减少这些生命中各自所受之苦的糟糕性。

[14]至于对人称性的价值和道德价值之间的相似区分，参见Nagel（3），第97—139页。

[15]考虑一下陀思妥耶夫斯基的小说《卡拉马佐夫兄弟》中大裁判官的那一席话，以及上面引述的R·B·佩里的主张："一百万人的幸福不知何故也完全不能……即便减轻一个人所受折磨的严重性。"

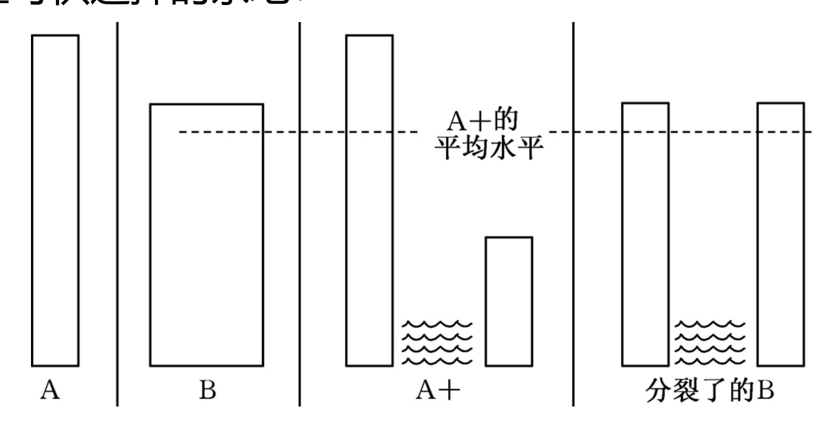
[16]按照词典式的观点，我们必须接受（A）的一个变体，以"平庸的"一词替代"没有价值的"一词。低于这个水平的10个10亿人的存在价值也比哪怕一个高于有福的水准的人的价值要小。即使这些人的存在比惟一的一个这样的人的存在具有较少的价值，但是其存在价值会多于一个受苦的、具有并不值得过的生活的人的存在所压倒的价值。我们还必须以同样的用词变化来接受（R）的一个变体。按照词典式的观点，当我们考虑高于平庸的水准之上的生活的时候，量总会压倒质。

第十九章

纯粹加法悖论

142.纯粹加法

考虑这些可供选择的余地：



假设我们诉诸那个没有价值的水准。假定B中的生活质量低于这个水准。那么我们将认为，尽管B中的人合起来具有比A中的人更多的幸福，而且具有更多别的什么使生活值得过的东西，但这并非是B比A更好的一个方面。B中的生活低于量具有价值的那个水准。我们还认为，质具有价值。我们认为，B中的所有人都比A中的所有人处境糟糕一些这一点是糟糕的。既然B在一个方面比A更糟，而没有什么方面更好，那么我们得出结论：B比A更糟。

假设我们转而诉诸那个词典式的观点。B中的生活质量大约相当于A中生活质量高度的4 / 5。鉴于这个事实，我们不能够令人信服地既假定B中的生活低于那个平庸的水准，又假定A中的生活高于那个有福的水准。照我们的观点，如果B足够大，它会比A更好。A在量上的充分增益会超过其低一些的质。但是既然B只有A的两倍大，我们能够主张这并不属实。我们能够再次得到B比A更糟的结论。

现在比较A和A＋。在A＋中，有一组人像A中的人一样大，而且具有一样高的生活质量。A＋包含另一组人，我称之为特大量的人。这些人具有值得过的生活，而且他们并不影响别的什么人。那些特大量的人处境比第一组人糟。如果这种不平等既是已知的又是可消除的，那么它可能牵涉到社会不公。我因而假定，出于简单性的考量，A＋中的两组人互相并不知道对方的存在，而且不能交流。A＋是人类横跨大西洋之前某种可能的世界状态。A则是一个与之不同的、美洲大陆无人居住的状态。

A＋比A好还是糟？界定一个新短语将有所帮助。当出现下述情况的时候存在

> 纯粹加法：当两个结果的一个结果中存在特大量的人，（1）他们具有值得过的生活，（2）他们并不影响到别的什么人，而且（3）其生存并不涉及到社会不公。

当我们比较A和A＋的时候，A＋中的纯粹加法降低了平均生活质量。这是造成A＋比A更糟的一个糟的后果吗？

143.我们为什么应当拒斥均量原则

根据均量原则，无论根据其对人产生影响的形式还是其无人称的形式，如果每个生活的平均生活质量较低一些的话，那就更糟。照这个原则，A＋比A更糟。但是我们应当拒斥这个原则。

照均量原则，最好的历史或许是亚当和夏娃所生活的那个历史。如果不是亚当和夏娃，而是数以十亿计的人生活的话，尽管他们有几乎同样高质量的生活质量，也会更糟。尽管这个主张难以令人相信，但是它并不荒谬。第二种历史在一个方面更糟。没有什么人的生活像亚当和夏娃会有的那样好是糟糕的。

均量原则有别的蕴含，它们是荒谬的。假设亚当和夏娃过着这些美好的生活。照均量原则，如果，不是代之以而是附加有，数以十亿

计的别的人生活的话，会更糟。这之所以会更糟，是因为会降低生活的平均质量。这个通过纯粹加法降低平均值的方法不可能令人信服地主张为糟糕的。

类似的主张适用于任何一个孩子的诞生。这是否糟糕，照均量原则，有赖于有关所有的先前生活的那些事实。如果古代埃及人享有非常高的生活质量，现在生一个孩子这一点糟糕的可能性更大。这个孩子的降生更可能会降低所过活的生活的平均质量。但是埃及学的研究与我们决定是否生养孩子并不相干。[1]

这些是对均量原则的那个时间上中立的版本的诘难。我们能够维护某个别的版本吗？一些学者假定，重要的是在我们的行动之后生活的那些人的平均生活质量。但是这荒谬地蕴含着，如果我们杀死除了现在生活着的处境最好的人之外的所有的人，会使结果更好。

可能主张说，重要的是所有现在生活着的人和未来的人的平均生活质量。这个主张避免了上面最后那个蕴含。但是考虑一下

何以只有法国幸存。在一个可能的未来中，世界上处境最糟的人们很快开始具有相当值得过的生活。不同国度的生活质量从而继续上升。尽管每一个国家都享有世界资源中它该享有的份额，但是像气候和文化传统等因素却带给某些国家更高一些的生活质量。处境最好的人在许多世纪是法国人。

在另一个可能的未来中，一种新的传染病使几乎所有的人都失去了生育能力。法国科学家生产出只够法国人使用的一种解药。其他国家的人都不再存活。这对幸存下来的法国人的生活质量产生一些糟糕的影响。于是，不再有什么外国艺术、文学或者技术可供法国进口。这些和其他一些糟糕的影响比任何好的影响更重大。在整个这第二个可能的未来中，法国人具有的生活质量因而略微低于第一种未来中可能有的那种生活质量。

在这第二种未来中平均生活质量会高一些。幸存的法国人会有更低的生活质量；而且在绝大多数人都失去生育能力的那段时期，这些人中的绝大多数人的生活会更糟。但是这两个影响会被世界上其他国家不再存在大大地超过。在第一种未来中，在这些国家中会有几十亿的人存在。在许多世纪，这几十亿人的处境会比法国人在无论这两个

未来的哪一个中都糟。如果这数十亿的人根本不存在的话，未来生活的平均质量从而会更高。

照均量原则，如果只有法国人幸存会更好。这是另一个荒谬的结论。如果这数十亿的人生活着的话，他们的生活是相当值得过的，而且他们的生存比法国人要更好。照均量原则，如果这些人生活着的话，之所以会更糟，只是因为法国人的生活要更好。由于此为真，这些人的生存会降低平均的生活质量。正如从前一样，降低这个均量的这种方式不可能令人信服地被认为是重要的。

一个类似的主张再次诉诸任何一个孩子的诞生。假定，在更遥远的未来中，生活质量在许多世纪都极端地高。那么，如果我生一个孩子，即便我的孩子的生活会是相当值得过的，而且他的存在不会对什么人来说是糟糕的，也更有可能是糟糕的。我的孩子的存在更有可能会降低所有未来的生活的平均质量。这不能是相关的。我是否应当生一个孩子不能取决于遥远的未来生活质量将是什么样子。

我们或许修正均量原则，以便它避免这个蕴含。这样的一个修正很可能会涉及到作出一个随意的区分。[2]而且那些主要的诘难仍然适用。均量原则必须涵盖相互重叠的生活。照这个原则，我是否应当生一个孩子仍然取决于有关其他人的生活的那些不相关的事实。

此点最为清晰地表现于

> 地狱丙。我们绝大多数人具有糟得无以复加的生活。只有造成我们受苦的那些虐待狂般的暴君是例外。我们中其余的人如果可能的话会了结自己的生命；但这并无可能。那些暴君由衷地主张，如果我们生孩子的话，他们将使这些孩子略微少受一些罪。

照均量原则，我们应当生这些孩子。这会提升平均的生活质量。我们的孩子们的生活会糟得无以复加并不相关。这是另一个荒谬的结论。

尽管还有其他一些对均量原则的诘难，但是我并不打算在此陈述。我们业已看到的足以让我们认识到，我们应当拒斥这个原则。

144.我们为什么应当拒斥诉诸不平等

再考虑一下A和A＋。A＋中的那些特大量的人具有值得过的生活，而且他们并不影响到什么别的人。A＋比A更糟吗？

特大量的那些人的生活降低生活的平均质量。但是，正如我所论证过的，这个事实是不相干的。当这个均值被纯粹加法降低的时候，我们不能够具有可信度地诉诸那个均量原则。

还有别的一个方面或许可以主张A＋比A更糟。在A＋中存在我所称的自然不平等。那些特大量的人比另一组人处境更糟，决不是出于他们的错。那些特大量的人并不知道这个事实，而且不存在社会不公。这个没有察觉到的自然不平等造成A＋比A更糟吗？

诘难者或许诉诸像罗尔斯所捍卫的那些正义原则。这些原则之一是

> 极小中的极大值：最好的结果是最糟的人的处境达到其最好值的结果。

“附录”H表明，极小中的极大值可能与罗尔斯的其他原则相冲突；但是我们在此可以忽略这一点。与罗尔斯不同，一些人把这个原则应用于所有情况。这些人或许主张A比A＋更好，因为A是那个其中最糟人的处境达到其最好值的结果。

假设，首先，使某人存在能够带来益处。照这个假设，如果A＋出现，这将有益于A＋中处境最糟的人们。A＋中存在不平等可能看起来是糟糕的。但是，那造成不平等的东西有益于所有那些比他人处境糟的人们。而且这是所有可能的结果中最有益于这些人的那个结果。我们在此能够令人信服地诉诸罗尔斯的另一个原则。照这个原则，如

果造成不平等的东西带给所有那些处境差一些的人们以最大可能的益处，那么不平等就不是糟糕的。这支持A＋并不比A更糟的主张。

假设，其次，使某人存在不能带来益处。照这个假设，如果A＋出现，这并不有益于A＋中的那些处境最糟的人。诉诸极小中的极大值会证明A＋比A更糟的主张吗？

正如我们在前面所论证的，理想契约主义方法不应当应用于那个应当有多少人存在的问题。极小中的极大值同样如此。在那些不同的结果中会有相同数目的人生存的情况下，假设我们接受这个原则。但是这并不能够推论出，我们应当把极小中的极大值展开到那些不同的结果中会有不同数目的人存在的情况。极小中的极大值在这两类情况下所蕴含的是非常不同的。

在一个具有两个结果的同样数目事例中，考虑这样的一个结果，其中

（1）那个最糟的群体处境达到最好。

如果我们把这个结果与另一个结果相比较，必定为真的是

（2）在这个结果中那些处境更好的人多于另一个结果中的那些处境最糟的人。

在那些同样数目事例中，（2）是（1）在两个结果之一的结果中能够为真的惟一的方式。如果（2）为假，处境最糟的那个群体在这两个结果中都会同等大小，同等糟糕。根据极小中的极大值，两个结果中更好的一个是其中（1）为真的那个。在那些同样数目事例中，当且仅当（2）为真的时候（1）才为真。因而我们能够主张，根据极小中的极大值，两个结果中更好的一个是其中（2）为真的那个。

现在比较一下A和A＋。如果我们在这个不同数目事例中诉诸极小中的极大值，它蕴含着A比A＋更好，因为A是其中（1）为真的那个结果。当在某个结果中只有一个群体的时候，这个群体既是处境最好的

群体，也是处境最糟的群体。在结果A中（2）为真吗？A中处境更好的人多于A＋中处境最糟的那个群体吗？不然。在那些不同数目的事例中，（1）在两个结果之一的结果中能够为真还有另一个方式。（1）能够为真，是因为

（3）在这个结果中某些人并不存在，而在另外的结果中他们会具有值得过的生活。

（3）不同于（2）。既然（1）以这种不同的方式为真，那么我们不能够简单地假定在那些不同数目事例中，诉诸极小中的极大值是得到正当证明的。既然（1）以这一不同的方式为真，这是否使结果更好，是一个公开的问题。如果（2）在两个结果的一个结果中为真，这显然是这个结果的一个好特征。如果处境更好一些的人比另一个结果中处境最糟的那个群体要多，这显然是好的。如果（3）在两个结果的一个结果中为真，这显然是个好特征吗？如果那些在另一结果中本会具有值得过的生活的人并不存在，显然是好的吗？这不能被主张为显然是好的。我认为，如果（3）为真，这不是一个好特征。（3）的真值并不使这个结果更好。因为（3）为真，所以（1）为真。处境最糟的那个群体处境更好。但是，如果这是（1）为真的方式，（1）便不是一个好的特征。如果处境最糟的那群人以这种方式更好的话，在道德上是不相关的。在这种不同数目事例中，我们不应当诉诸极小中的极大值。

某些人可能置疑这些主张。但是如果这些人在这类事例中诉诸极小中的极大值的话，他们所面对的诘难就像均量原则所面对的那些诘难。考虑一下只有法国幸存的那个事例的一个变体。假设，如果所有别的国家不再存在，这会大大地降低幸存下来的法国人的生活质量。但是法国仍然会比在没有什么国家不再存在的情况下处境最糟的国家更好。极小中的极大值因而蕴含着，如果只有法国幸存的话会更好。在这个事例的这个版本中，这个结论甚至更为荒谬。如果所有其他国家都不再存在，而幸存者们结果会更糟得多，如何能会更好一些？

既然它蕴含这个荒谬的结论，极小中的极大值不应当应用于这类不同数目事例。诉诸极小中的极大值不能够支持有关A＋比A更糟的主张。

一些平等主义者并不诉诸极小中的极大值。这些人可能主张：“既然你并不认为如果特大量的人生活会更好，那么你就不会认为有哪一方面A＋比A更好。而且你应当承认A＋在一个方面更糟。某些人并非出于本身的错而比其他人处境更糟，这是A＋的一个糟糕特征。既然A＋没有什么方面比A更好，反而在一个方面更糟，那么，A＋必定比A更糟。”

这可能看起来令人信服。但是对此能够以我回应先前的一个诘难的那种方式加以回应。其时我假定，既然世界的实际人口现在是巨大的，那么，如果值得过的特大量人口的生活得以进行，那也不会更好一些。对这个观点我提出了下述诘难：“比较一下世界实际人口与一个更大的可能人口。在这个更大的人口中，每一个人都享有一个高得多的生活质量，尽管每一生活都包含某种剧烈的受苦。每当你设想这个更大的人口甚至更大的时候，这个结果照你的观点没有什么方面会更好。但是这个结果在一个方面更糟，因为会有更多的剧烈受苦。既然这个结果没有什么方面更好，反而有一个方面更糟，你就必须同意它必定更糟。而且你必须同意，如果这个人口足够大的话，受苦的增加会超过这个结果的所有那些好的特征。尽管这个结果中所有的人会远远好于我们现在的状况，但是他们的生存会比世界实际人口的生存更糟。照你的观点，你不能够避免这个荒谬的结论。”

我提出了我们如何能避免这个结论。我同意，如果在某个结果中存在剧烈的受苦，这是一个糟糕的特征。但是我否认了每一避免这个结论的方式都会使结果更好。至少有两种方式会有更多的受苦。或许的确，要么（1）现存的人们受苦更多，要么（2）有特大量的人过着他们尽管值得过却包含某种受苦的生活。在这两种或许有更多受苦的

情况中，只有（1）使结果更糟。如果存在更多的受苦是因为（2）为真，存在更多的受苦这个事实并不使结果更糟。如果受苦更少一些是因为这些特大量的人并不存在，那就不会更好。只有当他们的确存在，而且受苦更少一些的时候，才会更好。

上文主张说，A＋中的不平等是一个糟糕的特征。我接受这个主张。但是我再次否定：每当避免这个糟糕特征的方式时都会使结果更好。不平等是否使结果更糟取决于它如何出现。可能的确，要么（3）某些现存的人变得比其他人处境更糟，要么（4）有特大量的人生活着，尽管他们的生活是值得过的，但是比某些现存的人处境要更糟。只有（3）使结果更糟。

当（4）为真的时候，不平等可能由我所称的纯粹加法产生出来。当有特大量的人生活着，他们享有值得过活的生活，他们并不影响到什么别的人，而且其生存并不涉及到社会不公的时候，就存在纯粹加法。当不平等由纯粹加法产生的时候，它并不使结果更糟。这种不平等在下面这些情况下将是可以避免的：要么（5）特大量的人存在而且不比别的任何人处境更糟，要么（6）特大量的人根本不存在。只有（5）会使结果更好。如果没有什么不平等是因为特大量的人并不存在，那就不会更好。只有当特大量的人的确存在，而且像别的任何一个人处境一样好的情况下，才会更好。

既然A＋中的不平等是由纯粹加法产生的，这个不平等并不使A＋比A更糟。我们不能令人信服地主张特大量的人本来根本不应存在过，只是因为，不为他们所知的是，存在着其他处境甚至更好的人们。

145.悖论的第一个版本

现在比较A＋与B。经由中间世界——分裂了的B，B人口的两半之间不能交流——作这一比较可能有所助益。显然B与分裂了的B一样

好。我们现在能够问：“如果A＋变为分裂了的B，那么，这会变得更好还是更糟？”

既然这两群人不能交流，这个变化不会是处心积虑的一种再分配的结果。从A＋到分裂了的B的改变会是影响到环境的那些自然事件的结果。而且这个改变可能慢慢地经历了两个世纪。在从A＋到分裂了的B的变迁中，更糟的一半的增益会多于更好的一半的损失。照我们平常的道德假设，这会变得更好。既然这是一个相同数目事例，我们能够诉诸极小中的极大值以及平等原则。根据这两个原则，分裂了的B都比A＋更好。还记得那个更糟群体之所以更糟并非出于他们自身的过错。而且我们能够假设，他们境况更糟，不是仅因为他们较不幸福，或者具有较低一些的生活质量，而是因为他们分享到更少的资源份额。我们能够假设，无论从福祉的平等还是从资源的平等的角度来说，分裂了的B都比A＋更好。（我在假定，在所有我所设想的事例中，生活质量、有福的水准和每一个人享有的资源份额都一同升降。）

判定分裂了的B比A＋更好还有一个根据。在某些事例中，平等与善行相冲突。当一个更糟的群体的获益少于一个更好的群体的损失的时候，情况就会如此。但是分裂了的B比A＋更好，既是按照任何平等原则又是按照任何具有可信度的善行原则来说的。更糟的群体的获益大于更好群体的损失。

或许说，那些平等原则只适用于可能存在社会不公的某个社会之内。如果是这样，分裂了的B按照任何一个善行原则仍然比A＋更好。这个有关平等的主张是不会令人信服的。考虑一下

贫与富。假设我认识两个人，他们生活在彼此不能沟通的不同社会中。我称为富者的人比我称为贫者的人处境更好。除非我去帮助富者，否则他的生活质量和资源占用份额就将出现下滑。我要么介入以保持富者的目前水准；要么代之以帮助贫者。如果我帮助贫者，我能够把他提高到富者在没有我帮助的情况下会滑落到的那个水准。而且贫者所提升的会大于富者会滑落的。

我们绝大多数人会认为，如果我帮助两者中的任何一个的话，帮助贫者而不是富者会更好一些。而且我们绝大多数人会认为，这会使结果更好，既因为贫者所获益的大于富者会失去的，又因为两者中没有哪一个比另一个更糟。即便是富者和贫者生活在两个（除了通过我）不能沟通的世界中，我们绝大多数人也会相信这一点。我们于是会认为，分裂了的B比A＋更好。既然B显然像分裂了的B一样好，那么B也比A＋更好。

假设我们认为A＋不比A更糟。我们现在认为，B比A＋更好。这些信念合起来蕴含着B并不比A更糟。如果B比某物——并不比A更糟的A＋——更好，那么，B就不可能比A更糟。但我们早先所相信的是，B是比A更糟。我们拥有三个不一致的信念，而且蕴含着矛盾。这些信念蕴含着，B既比A更糟又不比A更糟。我称之为纯粹加法悖论。[3]

这不只是不同道德原则之间的一种冲突。我们可能具有一种多元主义的道德，其中我们认为，如果既有更大的平等，又有更大的利益总量的话，会更好。然而可能存在更大的平等会降低利益总量的一些事例。我们的两个原则在此会发生冲突。但是在我们的道德观念中不会有什么不一致之处。我们可能只是要问，鉴于该事例的细节，平等方面增益的重要性是大于还是小于利益的损失。我们在此会力求达到一个把所有事情都考虑到的判断。在纯粹加法悖论中，那些事情是不同的。我们在此倾向于认为，把所有的事情都考虑到，B比A更糟，尽管B比A＋更好，而A＋并不比A更糟。对这三者不能连贯一致地都予以相信，因为它们蕴含着矛盾。这些信念中的一个必须放弃掉。

哪一个应当放弃掉？我们能够诚实地主张说我们认为，就A＋中的特大量群体而言，如果他们根本没有存在过的话，A＋本会更好吗？或者我们能够诚实地主张说我们认为从A＋到B的变迁不会是变得更好吗？如果我们主张后者的话，我们会是在说，最为重要的是境况最好的人们的生活质量。如果他们的生活质量滑落的话，即便同等大

的一个境况更糟的群体的生活质量的更大增益也不能在道德上超过这一点。即使境况更糟的人群之处境更糟并非出自他们自身的过错，也是如此。且称此为精英主义观点。照此观点，境况最好的人们发生什么比境况最糟的人们发生什么更重要。这个观点的一个更为极端的例子是极大中的极大值。这是与极小中的极大值对立的。照此观点，我们应当绝对优先保持或者提升境况最好的人们的生活质量。这两个观点都适用于现实世界。我们当中没有多少人会觉得在此它们道德上是可以接受的。

可能有这样的想法，即如果我们要么诉诸没有价值的水准，要么诉诸词典式的观点，那么我们就已经采纳了这个精英主义观点。情况并非如此。当我们考虑其中并不含有什么不平等的那些结果的时候，我们诉诸过那个没有价值的水准。假设A中的那些生活高于没有价值的水准，而B和Z中的生活低于这个水准。那么我们或许主张，A的生存具有内在的道德价值，而在B和Z的生存中不会有什么这样的价值。这可能就是B和Z为什么比A更糟的缘由。既然如果所出现的是A的话不会有什么不平等，那么这个主张可能会被平等主义者们所接受。

当我们比较B和A＋的时候，我们是在比较一些结果，其中一个里面含有不平等。照刚刚提出的那些主张，只有A＋中境况最好的那群人的生活拥有内在的道德价值。即便我们相信这一点，我们也无需主张A＋比B更好。我们的那些平等原则和善行原则蕴含着，B会比A＋更好。而我们有关内在价值的主张或许被这些别的原则所超过。有一个不同的却更带有精英主义意味的观点是，在并非出于自身过错的境况更糟的人群存在的一个世界中，最重要的是那群境况更好的人的生活质量。这个精英主义观点与我们的平等原则和善行原则相冲突，而且被这些原则所超过。

精英主义观点既不为那个词典式的观点也不为诉诸那个没有价值的水准所蕴含。但是可能存在这个两者的精英主义者版本。我们无需在所有的情况下都是精英主义者。我们可能只要主张，如果A中的生

活高于没有价值的水准，而B中的生活低于这个水准，那么，B就会比A＋更糟。照此观点，我们不会反对在那些境况更好的和那些境况更糟的人之间的所有再分配。我们会同意，在绝大多数情况下，对境况最好的那些人们来说的一种损失会被对那些境况更糟的人们来说的一种更大的获益所超过。只有当这样的再分配会造成境况最好的人们滑落到那个没有价值的水准之下的时候，我们才会反对这样的再分配。类似的主张适用于那个词典式的观点。照这个观点的精英主义者版本，只有当再分配会造成那些境况最好的人们跌落到那个有福的水准之下的时候，我们才会反对再分配。

假设，当我们比较B和A＋的时候，我们不能够接受精英主义观点的任何一个版本。我们认为，B比A＋更好。如果我们不能够认为A＋比A更糟，我们必须得出结论说，B并不比A更糟。我们必须得出结论说，如果这些是某个社会的两种可能的未来，如果所发生的是B的话——全都更糟的人们多达两倍的话，不会更糟。

纯粹加法悖论并不迫使我们接近这个结论。如果我们拒斥我们另两个信念中的一个信念，我们就能够避免该结论。尽管我们发现这些难以拒斥，但是我们发觉要接受B并不比A更糟这一点恐怕更加困难。假设我们断定，在两个避免这个结论的途径中，最不会令人难以置信的是A＋比A更糟。那么我们能够保持我们有关B比A更糟的观点。

不过，我们应当注意到，我们不能够简单地主张A＋必定比A更糟，因为它比某物——比A更糟的B——更糟。简单地基于它与其他两个之间不一致，我们在此本会拒斥这个彼此不一致的三个观点中的一个观点。针对每一个主张都可以这么说。但是为了避免此一悖论，我们必须只比较A和A＋，搁置起论证的其他部分，而且我们必须认为A＋更糟。我们必须认为，特大量的人生活着本身就是糟糕的。我们必须认为这是糟糕的，即便这些人享有值得过的生活，而且并不影响到

别的什么人。在我们发觉这令人难以置信的意义上，我们仍然面对着一个悖论。

可能有这样的诘难：“你的论证包含着一种圈套。当你比较A和A＋的时候，你主张特大量的人的生存不会对什么人而言将会更糟。但是当我们移到了B的时候，原初的那群人却变得境况更糟了。特大量的人的增加对原初的那群人来说是更糟的。这就是为什么A＋比A更糟的缘由。”

该论证可以重新陈述一下。假设我们在考虑许多世纪之前，比方说9世纪，世界的一些可能状态。我们没有什么根据担心未来的后果；我们知道后来所发生的。其次假设A＋是过去的这个世纪中世界的实际状况。那么我们可以问，如果实际状况是A的话，本来会更好吗？在问这个问题方面，我们能够假设A＋后来并未变为B。A＋中那些境况更糟的人的存在并未向更糟的方面影响境况更好的人。而且，既然那些群体不能够交流，就没有社会不公。鉴于这些事实，A＋会比A本来会有的情况更糟吗？境况更糟的那群人过活的话，本是糟糕的吗？我们还能够问另一个问题。事实上世界并未从A＋变为B。但是，如果它真变了的话，这本来会是向更好的一种变化吗？照该论证的这个版本，最后的那个诘难从根本上被削弱了。境况更糟的那群人的生存并非对境况更好的那群人来说更糟。既然如此，我们可能看起来被迫承认，A＋并非一个比A本来会有的事态更糟的事态。而且我们可能看起来被迫承认，从A＋到B的一种变化本来会是向更好的一种变化。从这两个主张推论出，B本来就不会比A更糟。[4]

对这个论证还有一个诘难。一些人说：“我们必须区分两种情况。如果A＋变成B本不可能，A＋就不比A更糟。如果这个变化本会可能的话，A＋是比A更糟。”[5]

在该论证的这个最近的版本中，我们能够添加有关从A＋到B的变化本来就不可能的假定。如果实际上的状况是A＋的话，A本来会更好

这点就令人难以置信。而且难以否定的是，如果从A＋到B的一个变化是可能的话，这本来会是一种改进。照刚刚给出的诘难，如果从A＋变为B本来是可能的话，即便这变化并未发生，我们也应当改变有关A和A＋的观点。如果一个并未发生的变化本来可能发生的话，如果特大量的人根本就不存在的话，本会更好。

如果我们刚才在讨论的是人们应当做什么，这样的一种主张可能是令人信服的。我是否应当以两种方式中的一种方式行动可能取决于对我而言以某个第三种方式行动是否会是可能的。但是当这样的一些主张应用于我在讨论的那一类的那些结果的时候，它们是不会令人信服的。这些不是对某个人或者某群人来说可能的某组行动的那些可以预见到的不同的结果。没有什么人或者人群选择实际的结果将是A、A＋还是B。我假设了A＋是过去某个世纪中世界的实际状况，而且假设了A＋事实上并未变为B。我然后问了，与A＋相比较，A是否本来会更好一些。这两个结果的相对的好不可能依赖于第三个结果是否将永远不发生，是否可能本来会发生。

现在可能出现这样的说法："假设这些结果算是那些不同的可能的行动的可以预见到的后果。如果我们问我们应当做什么，那么我们就消除该悖论。假设我们要么会导致A的产生，要么会导致A＋的产生，要么会导致B的产生。导致A＋的产生会是错误的。这之所以会是错误的，是因为有一个更好的结果B，这个结果B是我们本来可以导致产生的。但是导致B的产生也会是错误的，因为有一个更好的结果，那就是A。"

正如我们后来将看到的，该悖论本来能够涉及我们应当做什么。但是对这个诘难有一个更为简单的回应。它并不消除我们的悖论。它只是忽略它。任何一个悖论都能够忽略。但这决不是什么解决之道。

我们能够添加这些主张。我们的道德思维中的绝大部分可能关于我们应当做什么。但是我们也拥有一些有关不同的结果之相对好和相对糟的观点。正如我已经说过的，这些并非关于在应用于行动或者行

动者的意义上的道德的好或者糟。如果在一场地震中死了数以千计的人，它并非在这个意义上是在道德上糟糕的。但它是在具有道德相关性的一个意义上糟糕的。我们有关不同结果的好和糟的那些观点有时依赖于我们有关我们应当做什么的观点。但是这种依赖性时常颠倒过来。正如最后的那个诘难本身所表明的，我们的一些有关我们应当做什么的信念依赖于我们有关结果的相对的好的信念。既然后者这些信念形成我们的一些道德的基础，我们就不能够拒绝考虑一下有关这些信念的论证。我们不能够忽略A和A＋的相对的好，像这最后的一个诘难却忽略了这一点。这就是这个诘难为什么没有化解该悖论的缘由。

如果我们认为纯粹加法不能够使结果更糟，我们就会面对该悖论。一些人认为，纯粹加法使结果更好一些。他们主张A＋比A更好。这些人们也会接受我有关B比A＋更好的主张。这两个主张蕴含着B比A更好。

如果我们接受这些主张，而且拒斥精英主义观点，我们就不能够避免那个令人讨厌的结论。存在一个可能的结果C，C与B的关系就像B与A的关系。在C中有两倍的人，他们都比B中的每一个人境况更糟。上面的论证能够再度适用。如果我们断言B比A更好，我们必须断言C比B更好。按照同一论证，D会比C更好，E比D更好，而且沿着字母表可以依次类推。最好的结果会是Z：一个庞大的人口，其中所有的人都只具有勉强值得过的生活。

146.我们为什么尚未被迫接受那个令人讨厌的结论

可能看起来，即便我们只是主张A＋并不比A更糟一些，我们就得被迫接受那个令人讨厌的结论。可能看起来，如果B会比A＋更好的话，而A＋并不比A更糟，B必定比A更好。同理，C必定比B更好，D比C更好，等等。

这个推理假定，“不比……更糟”蕴含着“至少像……一样好”。这是一个自然而然的假定。但是，经过反思，在此这并未得到正当的证明。考虑一个像A＋一样的结果，但特大量的人享有某种程度的更高的生活质量这一点除外。称之为得到改进的A＋。这个结果明显比A＋更好。如果我们既相信A＋并不比A更糟，又认为得到改进的A＋比A＋更好，那么我们必须得出得到改进的A＋比A更好的结论吗？答案是否定的。我们可以主张，尽管得到改进的A＋比A＋更好，但这两者只是不比A更糟。[6]

既然“不比……更糟”并不蕴含“至少像……一样好”，那么最后这个主张是连贯一致的。而且在许多别的领域，这些就是我们所应当提出的那些类型的主张。考虑一下某个文学奖的三个候选人，其中一个是小说家，两个是诗人。我们可能主张，就那位小说家和第一位诗人来说，没有哪一个比另一个更糟。这不会是在主张，这两者不能加以比较。它会是在断言大致的可比性。有许多诗人会是比这个小说家更糟的候选人，也有许多小说家是比这第一位诗人更糟的候选人。我们是在主张，就这两者而言，有关他们各自的优点我们能够说出些重要的东西。两者中间没有哪一个比另一个更糟。他们属于同一类型。接着假设第二位诗人比第一位诗人略微好一些。（当我们比较两个诗人的时候，我们的判断能够更少一些粗略。）这个判断迫使我们要么得出第二位诗人比那位小说家更好，要么得出第一位诗人更糟的结论吗？并非如此。我们能够主张，尽管第二位诗人比第一位诗人更好，但是两者中没有哪一个比那位小说家更糟，而小说家也不比两者中的哪一个更糟。同样我们能够主张，就我所设想的世界的那些状态而言，得到改进的A＋比A＋更好，但是两者中没有哪一个比A更糟，而A也不比两者中的任何一个更糟。

粗略的可比性在某些情况下只是无知的结果。当此点属实的时候，我们认为存在有原则上的精确性或者完全的可比性。当我们把那位小说家和两个诗人中的任何一个加以比较的时候，如果惟一的那些

可能性是一个更好一些，或者两者确切地同等好，这一点就为真。在这样一种情况下，这点却并不令人信服。此处粗略的可比性是内在的，不是无知的结果。就普鲁斯特[7]和济慈而言，要么一个是更伟大的作家，要么两者确切地同等伟大，这一点为真吗？甚至在原则上都根本不会有这样的精确性。但是某些诗人是比某些小说家更伟大的作家，或多或少地更伟大一些。莎士比亚是比P·G·沃德豪斯伟大得多的作家，但是斯温伯恩[8]充其量也不会比他伟大得多。我认为，这种内在的粗略可比性，适用于某些种类的结果的好，以及一个人在道德上具有重大意义的方面是否境况比另一个人更糟这样的问题。[9]

当只有粗略的可比性的时候，“不比……更糟”不是一个传递性的关系。（关系R是传递性的，条件是，如果X与Y具有R关系，而且Y与Z具有R关系，X必定与Z具有R关系。）第一位诗人并不比那位小说家更糟，后者并不比第二位诗人更糟。这并不迫使我们改变我们有关第一位诗人是比第二位诗人更糟的观点。

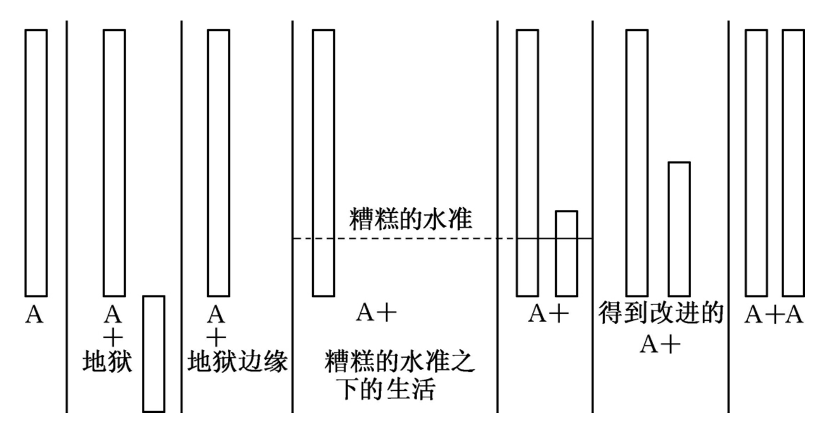
假设我们既认为B比A+更好，又认为A+并不比A更糟。因为“不比……更糟”此处并不蕴含“至少像……一样好”，我们并不被迫得出B比A更好的结论。我们只能够得出B并不比A更糟的结论。我们应该核实一点，就是在达到这个结论方面我们没有假定“不比……更糟”的传递性。我们没有如此假定。我们断定A+并不比A更糟，但比B却是更糟。我们得出了B并不能够比A更糟的结论。这个结论是得到了正当性证明的。假定相反的情况。假定B比A更糟。既然A+比B更糟，而且B比A更糟，A+必定比A更糟。这个论证是有效的，因为与“不比……更糟”不同的是，“比……更糟”是传递性的。但是我们拒斥这个论证的结论。我们认为，A+并不比A更糟。既然我们还认为A+是比B更糟，我们就必须拒斥这个论证的另一个前提。我们必定得出B不可能比A更糟的结论。

我们可能被迫到达这个结论。但是从此处我们甚至不可能被迫到达那个令人讨厌的结论的一个弱化的形式。的确，同理C不能够比B更

糟，D不能够比C更糟，等等。但是，既然“不比……更糟”并不是传递性的，我们就能够主张，尽管C并不比B更糟，后者并不比A更糟，但C是比A更糟。

147.诉诸糟糕的水准

对那个令人讨厌的结论还有一个更好的论证。在我陈述这个论证之前，先容许我讨论另一个观点。考虑一下下图所表明的那些结果的变动范围。



在A＋地狱中，那个特大量的群体具有（无罪）的生活，这些生活糟糕得无以复加。如果他们能够杀死他们自己的话，他们会的。显然，A＋地狱在道德相关的意义上是比A更糟。而且同样明显的是，A＋A并不比A更糟。在这两者之间的某处我们必须改变我们的观点。改变应当来自何处？某些人会说，就在特大量的人的生活变得值得过的那个水准上。照此观点，A＋地狱边缘并不比A更糟。

我们或许代之以采纳卡夫卡所提出的一个观点。他称某些种类的生活是受到限制的，而且主张，在其他事情同等的情况下，这样的生活“从道德观点来看是内在地不可取的”[10]。如果某个人过着一种受到限制的生活，倘若这个人根本从未存在过，而且没有什么人取代他的位置，这本身本来会更好。（既然它包含“本身”这个短语，所以这个主张并不涵盖对别的人们的影响。）

卡夫卡的观点的可信性取决于什么算作是受到限制的。当生活“在一般使人的生活具有价值和值得过的那些主要方面的一个或者多个方面出现重大匮乏”的时候，卡夫卡便称之为受到限制的。他补充说，这样的一些生活将是“总的来说典型地值得过”。当应用于那些相当值得过的生活的时候，他的观点会是令人信服的吗？

考虑一下“为人父母”这个方面，这是“一般使人的生活……值得过的那些主要方面”的一个方面。有一些人，尽管他们没有生育能力，但是享有相当值得过的生活。搁置对他人的影响不论，这样的一些人生活着这一点是糟糕的吗？答案是否定的。接着考虑一下一个严重的终身残疾。一些盲人具有相当值得过的生活。搁置对他人的影响不论，这样的一些人生活着这一点是糟糕的吗？如果他们根本从未存在过，而且没有什么人取代他们的位置，这本来会更好一些吗？答案再次是否定的。再考虑一下某个影响更严重的残疾。假设，因为某个人具有这样的残疾，这个人的生活相当值得过、甚或接近相当值得过这一点不为真。卡夫卡的观点在此更加令人信服。可能有一些人，他们的生活尽管值得过，但是如此有病或者如此匮乏，以至于即便不计对他人的影响，这些人生活着这一点也是糟糕的。

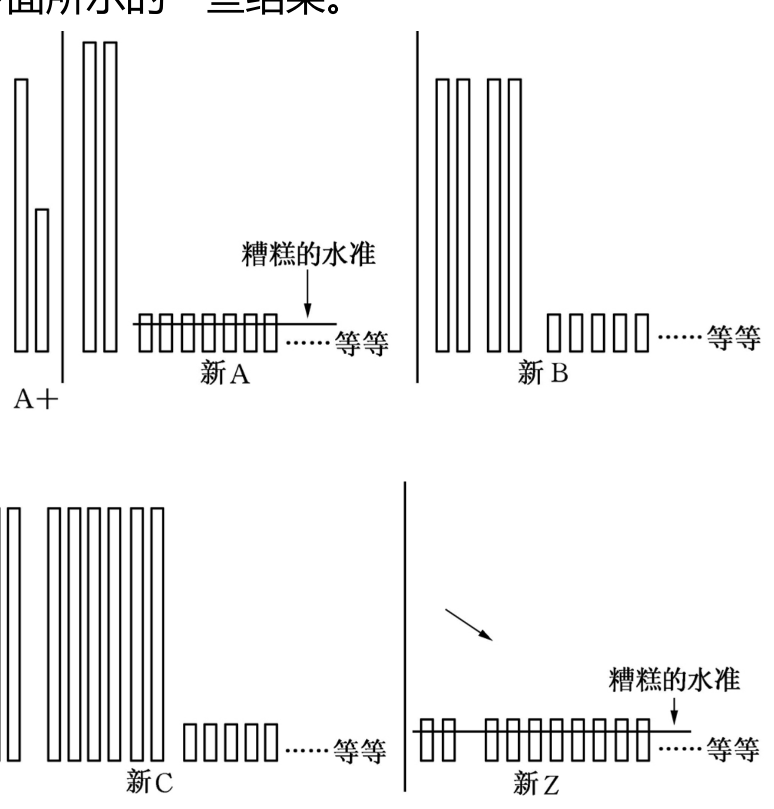
如果我们接受卡夫卡的观点，在这样的一些情况下，我们必须引入另一个水准，高于生活不再值得过的那个水准之上的一个水准。称之为糟糕的水准。我们现在可能主张，如果任何一个人在这个水准上或者低于这个水准生活的话，则是糟糕的。尽管这样的一种生活是值得过的，而且对于具有这个生活的那个人来说是有价值的，如果这个生活根本从来没有存在过的话，本来会更好。

卡夫卡的观点，为我们中间的某些人提供了对那个纯粹加法悖论的部分答案。我们可能认为，那个糟糕的水准高于生活不再值得过的那个水准之上。那么在像A＋中的特大量群体具有并不高于这个糟糕的水准的生活的那些情况下，我们将避免那个悖论。尽管这有所收获，但是收获不多。当应用于甚至接近相当值得过的那些生活的时

候，卡夫卡的观点是不能令人信服的。那些糟糕的生活必定比这更糟。它们必定在所有那些使生活值得过的特征方面是严重匮乏的。尽管它们值得过，但必定是不平顺的和不富足的。

148.悖论的第二个版本

考虑一下下面所示的一些结果。



每一个方块代表100亿人。这样A＋包含200亿人口。在A＋的这个版本中，即使境况更糟的那个群体也享有极端高的生活质量。

在新A中，存在着极端多的特大量人群。假定这些人群生活在别的一些太阳系中。新A是遥远的未来的一个结果。尽管这些人群都来自地球，它们现在却不能顺畅交流。

这些特大量人群中的所有人都具有比那个糟糕的水准高不了多少的生活。他们的生活是如此这般，以至于我们不能做到诚实地主张我们认为，如果他们根本从未存在过的话，这本身本来会更好。这点从

我对于糟糕的水准的定义而得出。因而我们不能认为新A比A＋更糟，因为这些特大量人们的存在本身是糟糕的。

新A至少有一个方面比A＋更好。在新A中有200亿人，其中所有的人都享有比A＋中的任何一个人都要高的生活质量。

新A中的不平等比A＋中的不平等更糟吗？我认为更好一些。在那两群境况最好的人之间不再有不平等。而仍然存在的不平等是由纯粹加法产生的。正如我所论证过的，当不平等是以这种方式产生出来的时候，它并不使结果更糟。因为这种不平等并不使结果更糟，而且在境况最好的群体之间不再有不平等，新A从平等主义者的角度看来比A＋更好。并没有可被称作使新A比A＋更糟的其他特征。既然新A在两个方面比A＋更好，而没有什么方面更糟，那么新A比A＋更好。

我有关不平等的那些主张或许遭到否定。按照某些观点，新A中的不平等比A＋中的不平等更糟。但是不能主张说糟得多。而且，即使有一个方面新A比A＋更糟，但还有别的方面新A更好：即境况最好的人们享有更高的生活质量这个事实。当我们比较这两个特征的时候，我们不能够令人信服地主张新A比A＋更糟。倘若我们否定新A更好一些，那么我们至少必须承认新A并不比A＋更糟。[11]

现在比较一下新A与新B。除了多出那些并不受到影响的群体之外，这就如同A＋和B之间的比较。因为存在有这些附加的群体，所以此处下面这一点并不为真：如果新A变为新B的话，这会消除自然的不平等。但是下面这样的一点会为真：尽管境况更好的那个群体会有若干损失，但是一个境况更糟的群体会得益若干倍。这个相对增益比早先论证中的增益大得多。

如同以往，这个变化不会是处心积虑的再分配造成的结果。它会以某种自然的方式出现，很有可能是因为环境的变化。除非我们接受极大中的极大值，或者精英主义观点的某个版本，否则我们不能够主张这个变化使结果更糟。那些境况更糟的人们大得多的增益必定比那些境况更好的人们小得多的损失更为重要。除非我们是精英主义者，

从而我们必须判定新B比新A更好。在仍然有某种不平等的那些情况下，存在关于哪一种不平等的模式更好或者更糟的数种观点。尽管这些观点在许多情况下并不一致，但是所有那些令人信服的观点无不一致认为，新B中的不平等比新A中的不平等糟糕性少。而且就像以前一样，新B并非只是从平等主义的话来说比新A更好。新B按照任何一个令人信服的善行原则来说都是更好一些。如果存在有从新A到新B的转变的话，那些境况更糟的群体所得到的会比境况更好的那些群体所失去的要多得非常多。

现在比较新B与新C。再一次，境况更好的那些群体会更糟。但是这些群体更糟的程度要比同样多的群体会更好的程度要小得多。同理，新C比新B更好。这样的推理一致沿用到新Z。在这个结果中，存在有某个巨大的人口，其生活比那个糟糕的水准高不了多少。新Z必定比新A更好，因为在新字母表上往下每顺延一步都被判为向更好的一种改变，而“比……更好”是传递性的。

另外不要忘记，正如我所论证的，新A比A＋更好。结合起来考虑，这些主张蕴含着

> 新的令人讨厌的结论：在两个可能的结果中的第一个结果中，会有两个100亿人的群体。其中的一个群体享有的生活质量远高于任何一个实际过的生活。尽管这个群体占有的资源份额大得多，但是这个群体不可避免地比另一个群体境况更糟。另一个群体享有的生活质量甚至更高。
>
> 在第二个可能的结果中，会有一个巨大的人口数目，其生活质量比那个糟糕的水准高不了多少。就这两个结果而言，第二个会更好。

一些人可能认为，新A并不比A＋更好，只是并不更糟。这些人必定接受这个新结论的一个弱化版本。他们必定主张，就A＋和新Z而言，第二个不会更糟。

这个新结论在一个方面比那个令人讨厌的结论较少令人讨厌一些。在Z中，人们的生活勉强值得过。在新Z中，人们的生活较之要更好一些。但是这个新结论在我看来非常令人讨厌。比那个糟糕的水准

高不了多少的生活不可能是相当值得过的，或者接近于相当值得过的。即便值得过，它们也一定被剥夺了绝大多数使生活值得过的东西。如果我们不能避免这个新结论，这就从根本上削弱了我们考虑人口过剩时我们绝大多数人所相信的观念。如果有200亿人，而所有的人都至少享有一个非常高的生活质量，那么我们会认为，这个结果比下述情况更好：如果代之以有更多的人，而所有人的生活尽管是值得过的，但却是被严重剥夺的、不平顺的和不富足的——比生活依其过活则本身会是糟糕的那个水准高不了多少。按照这个新结论，这些结果中的第一个结果会更糟。按照这个结论的那个弱化版本，第一个结果不会更好一些。

我们能够抵制这个新论证吗？或许有人提出，即便新B会比新A更好，而新C会比新B更好，但是这个推论并不会沿着新字母表一直适用下去。或许有主张说："当所有的群体都境况更糟的话，那么平等原则具有较小的权重。如果两个群体境况都好，那个境况更糟的群体的更大的增益在道德权重上超过那个境况更好的群体更小的损失。但是当平等原则应用于境况更糟的那些群体的时候，具有小的权重。而且还存在某个水准，在其之下这个原则便没有什么权重。"

这个观点并不令人信服。许多人认为，当平等原则应用于那些境况都更糟的群体的时候，它所具有的权重不同。但是这些人认为，在这样的一些情况下，该原则具有更大的权重。所提观点的对立面是他们所相信的观点。而且我不知道有什么人接受这个观点。[12]

此外我们应当记得，该论证并不只诉诸平等。沿着新字母表每往下推一步，境况更好的群体会失去的比同样多境况更糟的群体所获益的要少得非常多。无论是按照平等原则还是按照任何一个令人信服的善行原则，这都会是一个向更好一些的变化。除非我们是精英主义者，否则我们必定承认每一变化都是向更好一些的变化。

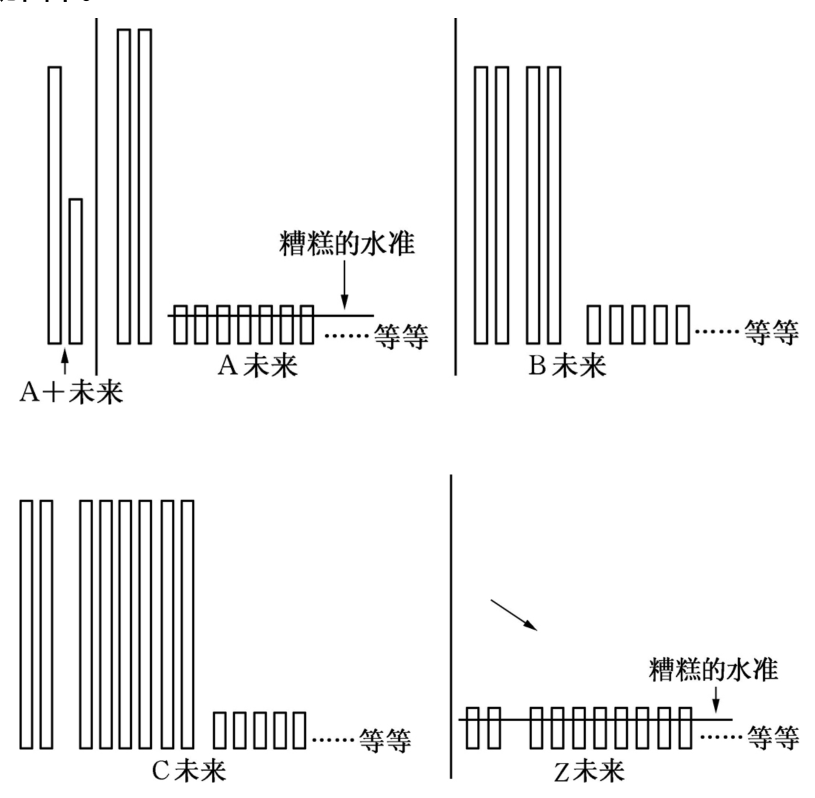
如果我们认为新B会比新A更好，我们就不能够令人信服地否定有关新字母表上那些依次往下毗邻的结果的可比主张。如果我们想要避免新的令人讨厌的结论的两个版本，我们从而要么必须主张新B并不比新A更好，要么必须主张新A比A＋更糟。正如我已经论证过的，除非我们能够证明精英主义观点的某个版本的合理性，否则这两个主张中就没有哪一个是可以得到辩护的。

现在我把论证概括一下。新A中的那些特大量的群体过着高于那个糟糕的水准的生活。鉴于我对这个水准的界定，我们不能够认为这样的一些生活得以过本身是糟糕的。这些人的生存并不影响什么人更糟。而且这些人的生存并不导入自然的不平等。我已经论证过，就平等方面而言，新A比A＋更好。那些持异议的人不能够主张，在这个方面，新A糟糕得多。而且在另一个方面，新A比A＋更好。多达200亿的人享有高于A＋中那些境况最好的人的生活质量。鉴于这些事实，不能够主张说新A比A＋更糟。我们也不能够诉诸有关新A与A＋相比可能具有更糟的后果的那个主张。我们可能主张说，新A中的群体如果更有交流能力的话，就将存在有应当予以消除的社会不公，再分配的结果会比A＋更糟。我已经解释了这个主张何以能够杜绝。我们能够假设，作为我们的事例的一个特征，这样的再分配将永远不会发生。实际结果将是，而且将一直是，新A。将永远不会有向新B或者新Z的变化。其他这些结果将仍然只是可能的结果。既然如此，我们只需比较新A和A＋的相对价值。如果特大量的人永远不存在，但代价是所有那些生活的人的将会更低的生活质量，这还会是更好一些吗？这令人难以置信。而且，要么认为新B本不会比新A更好，要么持有关于依次往下的两个毗邻结果的这样的一个信念，也令人难以置信。如果我们真的具有这样的一个信念，我们必定接受精英主义观点的某个版本。如果我们接受这个观点，我们必须把它应用于实际的世界。假设我们认为某些欧洲人的生活比那个有福的水准高不了多少。按照词典式的观

点的精英主义版本，某个损失如果对这些人来说会使他们降到那个有福的水准之下，这个损失在道德权重上不会被那些境况糟得多的人们——诸如那些患有营养不良的非洲儿童——远远大得多的一种增益所超过。这一点难以相信。但是新Z比A＋更好也难以相信。我们的难题仍然没有解决。

149.第三个版本

我们需要考虑那些在不同的结果中包含着所有会存在的人的情形。上面最后那个图解能够表明这样的一种情况。它能够成为如下面的图形所示的那样。



这个图形表示一些可能的不同的未来。现在每一个方块代表人类未来历史中的1000年。方块的高度表明在这1000年中每一个人享有的生活质量。在可能的这些未来的所有时刻都有100亿人生活着。

A＋未来于23世纪开始。在这之前的两个世纪一切安好。当A＋未来开始的时候，不再有不同人之间的不平等，而且生活质量甚高。前后相继的一代又一代的人享有这个生活质量达1000年。然后太阳变得比以前要热得多。这造成生活质量在许多方面低了许多。即使在这样的情况下，甚高的水准又保持了1000年。然后太阳变得非常炎热，从而终结了人类历史。

在A未来中，前2000年过得更好一些。每一个人都享有高于A＋未来中境况最好的人们的生活质量。在这期间太阳没有变化。在这个阶段末期，科学家们预言太阳将变得炎热得多。人们因而挖掘了许多深深的洞穴。这些洞穴使人类躲过地球表面的炽热而幸存下来。人们在这些洞穴中生活了许多个1000年。地下的生活是值得过的，但是远不如地球表面上一直拥有的那样好。在洞穴中生活的那些年中，生活质量并不比那个糟糕的水准高多少。然后太阳爆炸，终结了人类。

我们应当假定，A未来是实际上所发生的。其他那些未来只是一些可能的选择余地。与A未来相比A＋未来会更好吗？如果没有挖掘那些洞穴，结果历史在最初的两个1000年之后终结，这会更好一些吗？在这2000年期间甚至以较低的生活水准为代价这也会更好吗？

那些洞穴中的人们拥有值得过的生活，而且高于我们的糟糕的水准之上。令人难以置信这些人的生存本身是糟糕的。而且这比该论证的那个先前版本更加难以相信。在A未来中，在那些生活在任何一个时期的人们之间没有不平等；只有在一代人与另一代人之间才有不平等。这点强化了这样一个主张，即既然这个不平等是由纯粹加法产生的，那么它并不使结果更糟。

令人难以置信：如果挖掘了那些洞穴，结果特大量的人得以生存，这一点是糟糕的。纵然这是糟糕的，这也不会表明A未来比A＋未来更糟。这些未来在另一个方面有差异。如果我们认为A未来更糟，我们必须认为，如果特大量的人根本不存在，以凡将生活的每一个人的更低的生活质量为代价，则更好一些，这个信念是荒谬的。

在该论证的这个版本中，我们可以问人们应该做什么。考虑一下通过挖掘洞穴而逃过死亡的那些人。如果我们认为，A未来比A＋未来更糟，我们大概将会得出这些人不应当生育子女的结论。而且我们必须得出结论认为，他们具有不生育子女的道德理由，因为他们生养子女会使结果更糟。这并不令人信服。如果这些人具有子女的话，这些子女以及他们的后代都会享有值得过的生活。鉴于这个事实，令人难以置信这些人会有任何不生育子女的道德理由。这印证了我有关A未来并不比A＋未来更糟的主张。而且，在另一个方面，A未来将会更好。在最先的两千年中，每一个人都会享有一个更高一些的生活质量。既然A未来并不会在什么方面比A＋未来更糟，反而在一个方面更好，我们应当得出结论，就这两者而言A未来会更好。

现在考虑一下B未来。它在下面这个方面不同于A未来。在最先的2000年，其时在地球表面生活，生活质量会略微低一些。但是在随后身居洞穴的2000年中，生活质量会有远远大得多的提升。正如图解所示的，尽管生活在地表和地穴中会有不同，生活质量在最先的4000年中会是相同的。

A未来将是实际上所发生的。B未来会更好吗？境况最好的人们在最先的2000年中会有更低的生活质量。但是他们所损失的比相同数量的人们在随后的2000年中所获得的要小得多。按照我们通常的平等原则和善行原则，这会是一个更好的结果。

如上所述，这两个差异可能会由环境中的那些自然变化所产生。但它们也可能会由深思熟虑的行动所产生。假设太阳的变化在最先的2000年之初就预测到了。那些生活在这些年份中的人们，或许能够以自身的某些代价赋予那些在随后的2000年中生活的人们大得多的利益。一些人会主张，既然这无论按照平等原则还是按照善行原则都会使结果更好，这就是这些人们所应该做的。另一些人可能主张，这样的利他主义不会是一种义务，只不过是在道德上可赞扬的而已。这两者中无论哪一个都支持我的论证。为了反对该论证，我们会不得不主

张这样的利他主义不会使结果更好。如上所述，如果我们认为这个结果不会更好，我们必须接受某种形式的精英主义观点。我们必须认为，发生在那些境况最好的人们身上的事情比发生在那些境况更糟的人们身上的事情更为重要一些。境况最好的人所遭受的损失在道德权重上不会被境况更糟的人们大得多的增益超过。我们绝大多数人会发觉相信这一点是不可能的。那么我们必须承认，B未来比A未来会更好。

同样的评述适用于C未来。这会以类似的方式区别于B未来。最先4000年中的人们会出现损失，但是在紧接着的4000年中的人们会有大得相当多的增益。按照平等原则和善行原则，C未来会更好。同理，D未来会更好，E未来会更好，依次类推。所有这些可能未来中最好的会是Z未来。在人类历史的整个其余部分，生活质量并不比A未来中那些洞穴里会有的高多少。生活质量总归接近那个糟糕的水准。

既然"比……更好"是传递性的关系，那么Z未来会比A未来更好。而且正如我所论证的，A未来会比A＋未来更好。于是我们必定得出结论认为，Z未来会比A＋未来更好。与数以10亿计的人都至少享有一种甚高的生活质量的一个未来相比，倘若代之以存在有更多一些的人，会更好：他们的生活都不比我们借以认定这些生活若得以过会是糟糕的那个水准高不了多少。

在这个版本中，该论证更强一些。而且我们并不回避有关人们应当做什么的那些问题。我们问及这些问题，而且那些回答支持该论证。如果我们发现同样难以相信这个论证的结论，那么该悖论更大一些。[13]

---

[1]在此处我受益于McMahan（1）。

[2]再次参见McMahan（1）。

[3]本章第一部分重复了Parfit（7）的观点，而第二部分表达了一个新论证。

[4]F.Myrna Kamm and J.L.Mackie两人都提出，尽管按照平等原则把A+变为B会是我们的义务，但是这个改变不会是一种改善。我们可能有义务做或许使结果更糟的事。这个观点或

许提供了对于纯粹加法悖论的一个局部解决的方法。但它不会是一个完全的解决方法。

[5]这个诘难是由Tooley，Woodford和其他一些人所提出的。

[6]我把此点归功于R.M.Dworkin和A.K.Sen。我和其他很多人10多年来忽视了这一点。

[7]普鲁斯特（Marcel Proust，1871—1922），法国小说家，意识流小说的先驱。长篇小说《追忆逝水年华》是其代表作。小说强调“直感”，着重“潜意识”表述，对欧美现代派文学产生了相当大的影响。——译者

[8]斯温伯恩（Algernon Charles Swinburne，1837—1909），英国诗人。主要作品有诗剧《阿塔兰忒在卡吕冬》，诗集《诗与谣》、《日出前的歌》等。——译者

[9]参见Sen（1）中对于部分兼容性的讨论。

[10]载Kavka（4）。

[11]某个变化是否使某个不平等更好或更糟时常并不清楚。这个问题在很有可能会公开出版的Temkin的著作中得到很好讨论。

[12]再次参见Temkin。

[13]至于这个论证的一个修正版本，和一些进一步的思想，参见我的《人口过剩与生活质量》，载《实践伦理学》，P·辛格编，牛津大学出版社，1986年。

第二十章

结语

当人们问及西季威克的著作的时候，他说，他的著作的第一个词是伦理学，而其最后一个词是失败。这或许也是我这本拙作的“第四编”的最后一个词。正如我所论证过的，我们需要一个有关善行的新理论。这个理论必须解决那个非同一性问题，必须避免那个令人讨厌的结论和那个荒谬的结论，必须化解那个纯粹加法悖论。我未能寻获一个能够满足这四个要求的理论。尽管我未能寻获这样的一个理论，但是我相信，如果别人加以尝试的话，或许能够成功做到。

在本书的其他一些部分，我达成了各种各样的结论。它们绝大多数具有一个共同的特征。

150.非人称性

我的两个论题是理与人。我已经论证了，我们据以行动的那些理由应当变得更为非人称性。更大程度的非人称性可能看起来令人畏惧。但是它对每一个人来说时常会更好一些。

第三章论证了，在关心他人方面，我们大多数人犯了一些错误。我们大多数人都想避免伤害到他人。但是许多人认为

> （第二个错误）如果某个行动的对错归因于其后果，那么惟一相关的后果就是这个特定行动的那些后果。

这导致这些人忽略了它们合起来所做的。而且我们绝大多数人认为

> （第四和第五个错误）如果某个行动对别人的后果微不足道或者感觉不到，那么这个行动不能因为具有这些后果在道德上就是错误的。

这些虚假的信念在一些小共同体中并不重要，绝大多数人在绝大多数历史中就生活在这样的共同体中。在这些共同体中，只有当我们每一个人实质性地伤害到的那些人们存在的时候，我们才伤害到他人。

我们绝大多数人现在生活在一些大共同体中。我们行动的糟糕后果现在能够殃及数以千计，甚至数以百万计的人们。我们的虚假信念现在是严重的错误。在我所设想的那个无害的施虐者们事例中，它们是一些错误这一点非常明显。这些施虐者中的每一个都明知故犯地但不为人察觉地影响到1000个受害者中的每一个人所遭受的痛苦。这些施虐者的行动是非常不对的。他们知道，尽管他们中没有人造成任何可以察觉的差异，但是他们合起来施加给他们的受害者非常严重的痛苦。

这类实际事例数不胜数。在这些事例中，的确，每一个的行动对他人的影响微不足道或者无法察觉。我们错误地认为，因为的确如此，我们的行动后果不能够使它们成为错误的。尽管我们的每一个行动具有微不足道的后果，但是，我们合起来却极大地伤害到我们自己或者他人这一点却时常是真的。这方面的一些例子有：污染、拥挤、损耗、通货膨胀、失业、经济衰退、过度捕捞、过度耕种、水土流失、食物匮乏和人口过剩。

我们存在这些错误的信念，我们的无知是一个借口。但是在我们看清这些信念之错误的时候，我们就不再有什么借口。如果我们仍然继续以这些方式行动，那么我们的行动在道德上就将是错误的。某些可能像无害的施虐者们的行动那样糟糕。

合理的利他主义者们并不持有这些错误的信念。如果我们全都是合理的利他主义者的话，对每一个人而言这可能会更好。但是合理的利他主义者在这个意义上是更加非人称性的：他们并不简单地问"我的行动对任何一个人而言将更糟吗？将会有任何人抱怨吗？"他们认为，他们的行动是否可以察觉地伤害到他人并不相关。

大城市生活令人不安地非人称性。除非我们采用以毒攻毒之策，否则我们不能够解决这个难题。正如我们需要小偷来抓小偷一样，我们需要那些非人称性原则来避免非人称性的糟糕后果。

第四章论证了，因为常识道德时常是直接地集体性地自败的，所以必须予以修正。修正过的版本R是部分地向合理的利他主义的一个移动。我们必定再一次更加非人称性。

考虑一下我们对子女的义务。根据常识道德，我们应该给予我们自己的子女某种优先权。根据R，在一些情况下我们不应该给予我们自己的子女这些种优先权。我们所做的，不偏不倚地考虑下来，应该对每一个人的子女来说都会是最好的。通过告诫我们忽略与我们自己子女的关系，R告诉我们要忽略我们的人际关系中那个可能最强的关系。

如果我们都遵循这个非人称性原则，而且并不给予我们自己的子女任何优先权，那么这对我们的所有子女而言都将更好。非人称性再次显得更好，即便从人称的角度来说也是如此。类似的主张适用于我们与这样一些人之间的关系，诸如我们的父母、朋友、邻居、学生和病人等等。

第二编论证了我们应当拒斥那个有关合理性的自利论。S是最为强调人们之间差异，或者说个人的独立性的理论。S告诉我要做对我来说最好的事情。对S而言，基本的单位是不同的生活。我的至高无上的关涉是我的整个生活尽可能好地过。每一个人都合理地被要求给予他自己和他自己的生活绝对的优先权。

既然我们应当拒斥S，我们的理论必须在一个方面更加非人称性一些。千万不能主张每一个人至高无上的关涉应当是他自己；千万不能赋予生活之间的界线至高无上的重要性。但是我们的理论无需是西季威克的不偏不倚的仁爱原则。我们应当接受批判性的当前目标论，

或曰CP。按照这个理论，基本单位不是贯穿他的整个生活的那个行动者，而是正在行动的那一刻的行动者。尽管CP否定自利之至高无上的重要性和个人之整个生活的重要性，但它并不是非人称性的。CP主张，对我来说现在所做的合理的东西是什么，有赖于我现在想要什么、重视什么和相信什么。这个主张赋予每一个人的特定价值或信念以更多重要性。由于CP赋予使不同的人们区分开来的东西更多重要性，因此在这个不同的方面它比S具有更多人称性。

第三编论证了另一种非人称性。当我们考虑各种各样的想象出来的情形的时候，我们揭示出我们认为自己所是的东西。我们绝大多数人认为，我们的同一性是确定的。我们认为，对"我将要死了吗？"这个问题，总有一个必定要么肯定要么否定的简单答案。

除非我们是有别于我们的大脑、躯体和经验的独立存在的实体，否则这个信念就不可能为真。这些实体的连续存在必定是有别于物理的和心理的连续性的一个深刻的进一步事实，而且是一个要么完全成立，要么一点也不成立的事实。一个这样的实体会是笛卡儿式的自我。正如我们对那些所想象出来的事例的反应所表明的，我们绝大多数人倾向于认为我们是这样的一些实体。正如我所论证的，这并不为真。

因为这并不为真，所以我们不能够凭借主张这个生活中的那些经验全为这个人所拥有来解释个人生活的统一性。只有通过表述这些不同的经验之间所具有的各种各样的关系，以及它们与某个特定大脑之间的关系，我们才能够解释这个统一性。我们因而会以一种非人称性的方式来表述一个人的生活，而这种方式并不主张这个人存在。

按照这个还原论的观点，一个个的人并不存在。但是他们仅以国家存在的那种方式存在。个人并不像我们错误认为的那样是基本的。这个观点在这个意义上更加具有非人称性。

这个观点支持那些既关涉合理性又关涉道德性这两者的主张。按照那些极端主张，那些蕴含是完全非人称性的。有一位学者主张，如果还原论的观点是真的，那么我们应当把生死置之度外。另一些人主张，我们就不会有什么理由对我们自己的未来特别关注，而且绝大部分道德会从根本上遭到削弱。这些学者们认为，只有个人同一性的那个深层的进一步事实才赋予我们特别关心的一些理由，并且支持绝大部分道德。好像没有什么论证驳倒这个观点。因而提出这样的主张是可以得到辩护的：既然没有什么这样的进一步事实，那么我们没有什么特别关心的理由，而且道德的绝大部分从根本上遭到削弱。

尽管这些极端主张是可以得到辩护的，但是它们也是能够得到辩护地予以否定的。我提出了下述较不极端的主张：

> 心理联系的弱化可能既降低过去罪过的责任，又降低履行过去承诺的义务。
>
> 按照还原论的观点，拒斥那些分配原则是更加令人信服的。更加令人信服的并非聚焦于个人而是聚焦于经验，并主张道德上重要的是这些经验的本质。按照非人称性的功利主义原则，谁具有经验的问题就像该经验什么时候得以具有的问题一样毫不相干。这个原则忽略生活之间的界线，或者个人之间的独立性。按照还原论的观点，这个原则更加可信一些。（我并不是指"比其否定更加令人信服"。我是指"比按照非还原论的观点所会是的更加令人信服"。）
>
> 我们应当相信，极大的轻率在道德上是错误的。这减低了那些有关个人自主性的主张。当我们只影响到我们自己的时候，我们也不再具有随心所欲地行事的权利。在没有什么好的理由的情况下强加给自己巨大的伤害是错误的。

这些主张再次既降低了每一个生活的同一性的重要性又降低了生活之间界线的重要性。如前所述，我的那些结论是更加非人称性的。

有两个例外。一些学者主张，幸存中所重要的是物理的连续性，或者同一特定大脑的连续存在。这些学者们主张，如果我将被电子传输的话，我应当把这个前景当作几乎同死亡一样糟糕。尽管我的复制

品会与我在心理上完全相连续，但这不是所重要的。所重要的是他不会在物理上与我相联系。我不予赞同。我认为，所重要的是关系R，心理的连续性和／或联系性。通过论证这才是重要的，而物理的连续性并不重要，我是再次抨击在一个方面更带非人称性的一个观点。按照这个观点，所重要的是一个我们既与纯粹的动物分享又与纯粹的物体分享的特征。按照我的观点，所重要的是使我们之所以为人的东西。

我现在要增加一个类似的主张。按照非还原论的观点，每一个生活的深度统一性是自动确保的，无论这个生活是如何随机地、短视地和被动地去过的。按照还原论的观点，我们生活的统一性是一个程度问题，而且是某种我们可以影响到的东西。我们可能想要我们的生活具有更大的统一性，就像艺术家想要创作出一个具有统一性的作品一样。而且我们能够给予我们的生活更大的统一性，就以那些表达或者实现我们特定价值和信念的方式。既然还原论的观点更加看重我们选择如何生活，而且更加看重使不同的人们区分开来的那种东西，这是它更加人称性的第二个方面。

在第四编，我的那些结论是在最明显和最强烈的意义上非人称性的。如果我们想要避免那个令人讨厌的结论，我们就不能够凭借诉诸一个对人产生影响的原则来化解那个非同一性问题。我们必须诉诸这样一个原则，这个原则是关于所过的那些生活的质与量的，而不是对于我们的行动所影响到的那些人们而言是好与坏的。

我还论证了，如果我们诉诸这样的一个原则，那么对人产生影响的原则所蕴含的是什么这一点并不造成什么道德上的差异。当我考虑那两个医疗方案的时候，我得出了这样一个结论，即只是撤销这些方案中的一个方案对受到影响的孩子们会更糟这一点并不相关。如果我们必须诉诸这样的一个原则，而且从而应当忽略那些对人产生影响的原则，那么这具有广泛的理论蕴含，而且是带有非人称性的那一类。

151.论证的不同种类

伦理学追问结果是好还是坏，追问行动是对还是错。元伦理学追问道德语言的意义或追问道德推理的本质。后者还追问伦理学是否能够客观——是否能够得出为真的主张。

一些人认为伦理学只有两种治学方式，或者只有两种有关道德论证的方式。一种为低路线，即只是诉诸我们的直觉。另一种为高路线，即元伦理学。如果我们能够最好地说明道德推理的性质，我们就能够指望这会蕴含关于道德的特定主张。我们能够指望我们的元伦理学将会蕴含伦理学中的那些结论。

我认为，这些并非惟一的一些我们能够进行道德论证方式。除了假定一个可以接受的道德理论不能是直接地集体性地自败的之外，我并未采用高路线。我经常采用的是低路线，诉诸我们的直觉。但是我的主要目标之一是探索各种各样不同种类的论证，它们处于低端和高端之间。

第一章和第四章讨论

（a）一个理论是自败的论证。

以这种论证，我们能够取得无可否认的进步。既然S和C能够是间接地自败的，它们两者必定得出有关我们的愿望和性情的新主张。既然常识道德时常是直接地集体性地自败的，那么，按照几乎每一个元伦理学的理论，这蕴含着，这种道德必须得到修正。

第三章诉诸

（b）其道德意义一直被忽视的那些事实。

一个这样的事实是一组组行动的综合后果，或我们合起来所做的事的综合后果。另一个这样的事实是那些个别地来看无法察觉影响的综合后果。因为他们的行动后果之故，我所想象出来的那些无害的施虐者们行动非常错误，尽管没有任何一个施虐者造成任何一个受害者的痛苦可觉察到地更糟一些。这驳斥了这样一种观点，即一个行动不能因为其对他人的后果而是错误的，如果这些他人根本不会注意到任何差异的话。

第十六章诉诸同样类型的论证。它诉诸我们能够轻易影响到未来人的同一性这个事实。这个事实的蕴含——我所称的那个非同一性问题——对我们的那些道德理论来说具有明显的重要意义。但是，除了一些例外情况，我们一直忽略这些蕴含。

这些就是这种论证的惟一一些例证还不太可能。可能还有许多尚未发现的此类论证——还有许多别的事实的明显合理的意义或道德的意义只是被忽视了。这是另一种我们能够取得无法否认的进步的论证。

第六章诉诸

（c）对一个理论之假定和蕴含的更加完全的表述。

我针对自利论的第一个论证比一个问题强不了多少。但是要触及到这个问题，我不得不引入那个批判性的当前目标论。我的目标是孤立S，以便该理论能够以其自身的长短接受判定。因此我需要对照S和CP。诉诸工具性的理论或协商性的理论是不够的。这两个熟悉理论中没有哪一个以自利论者所不能忽视的方式挑战S。CP提供这样的一种挑战。而且它能够使我们看清S蕴含地所假定的东西。我们能够追问CP的哪一个版本会与S相契合。这个版本主张，自己所喜欢的一个时间上中立的偏向是再合理不过了。它主张，这个偏向必定是我们主导性的关涉，即便是在明知这些事实和思考清楚的情况下我们既不具有

也不想具有这个偏向。当我们更加清楚地看清S所假定的东西的时候，它就不再令人信服了。

第七章诉诸

（d）理论结构中的弱点。

S是一个混合理论，因为它要求时间上的中立性，却拒斥不同人之间的中立性要求。这并未使S不连贯一致。但这是一个结构性的缺陷，使S容易受到左右夹击。[1]

第八章提及另一种论证。这种论证诉诸

（e）形而上学的结论——关于实在或宇宙的最基本特征的结论——的一些蕴含。

一些哲学家和物理学家主张，时间的流逝是一种幻觉。如果我充分地辩护了这个主张的话，我本来会论证时间的中立性不可能是非理性的。

第三编诉诸这种类型的一个论证。我主张说，我们绝大多数人持有一些有关我们自身性质和我们历时的连续存在的性质的错误信念。如果我们能够表明我们具有这样的一些错误信念，那么诉诸真理可能支持某些既关涉合理性又关涉道德性的主张。于是，正如我主张的，还原论的观点提供针对自利论的另一个论证。而且这个观点支持各种各样的道德主张。

第四编诉诸那些（c）类论证。古典功利主义蕴含那个令人讨厌的结论，而均量原则蕴含一些荒谬的结论。第四编还诉诸一个（d）类论证。这个论证驳倒了任何一个时期量的价值都有一个上限的观点。

当这个观点展开到涵盖未得补偿的受苦的时候，它具有一个结构性的缺陷。像S一样，它是一个混合观点。尽管它主张说，量的肯定价值有一个上限，但它却不能具有令人信服地设置量的否定价值的一个限度。因而它蕴含另一个荒谬结论。

152.对我的那些结论我们应当欢迎还是痛惜

我论证道

（1）既然自利论和后果论时常间接地自败，那么它们必定对我们的愿望和意向有所主张。它们必定主张，我们倾向于以它们认定为非理性的和道德上错误的那些方式行动。

（2）既然常识道德时常直接地集体性地自败，它必须得到修正。

正如我在第五章中所提出的，这两个结论降低了常识道德和后果论之间的分歧。这是一个值得欢迎的结果。它把我们指向一种会消除这个分歧的统一理论。

我论证道

（3）在考虑我们的行动如何影响到他人方面，我们绝大多数人都犯了严重错误。因此，我们应当改变我们现在的许多行动方式。

这又是一个值得欢迎的结论。当我们看清这些是错误的时候，我们将更有可能以对我们大家都更好的方式行动。

我论证道

（4）我们应当拒斥那个关于合理性的自利论，接受批判性的当前目标论。按照这个理论，某些愿望是非理性的，而其他一些可能是合理地要求的。假设我知晓这些事实，在清楚地思考，而且我的那一组愿望并不是非理性的。如果会造成那个时候我最想要或最珍视的东西落空，那么对我来说以最符合我自己利益的方式行动会是非理性的。

（5）因为我们应当拒斥那个自利论，我们应当主张极大的轻率在道德上是错误的。

这些结论难以评估。我认为，原则上，结论（4）又是一个值得欢迎的结论。至少存在两种自利行动：

（i）这些行动中的某些行动有益于行动者，但带给他人更大的负担。按照自利论，如果不以这种方式行动是非理性的。如果人们不再持有这个信念，以这种方式行动的人将更加微乎其微。这会使结果更好。

（ii）某些自利的行动大大有益于行动者，而且对他人而言并非更糟。那么，如果更少的人以这种方式行动就是糟糕的。未能以这种方式行动是极大的轻率。这总归是糟糕的，而且通常是悲剧性的。我论证过，我们应当展开我们的道德理论，以便它主张极大的轻率道德上是错误的。如果我们接受结论（4）和（5）这两者，这有可能非但不增加反而会减少巨大轻率的发生率。但是可能有一些人接受结论（4）却拒斥（5）。如果人们不再认为极大的轻率是非理性的，而且继续认为它在道德上不可能是错误的，这就可能具有糟糕的后果。而且有一些人，对他们来说"非理性"这一指控具有比"非道德"这一指控更大的权重。

拒斥自利论在另一个方面也是值得欢迎的。与CP相比，S对道德的威胁更大。S与道德相冲突的情形有许多。无论我们想要什么和珍视什么，这些冲突中的一些冲突是不可避免的。如果我们相信S，那么我们认为在这些情形中道德地行动会是非理性的。这个信念可能造成我们更不大可能道德地行动。

类似的主张并不适用于CP。按照CP的一些版本，CP与道德相冲突的情形有许多。但是这种冲突并不是不可避免的。如果我们足够在意道德的话，该种冲突就会消失。

我论证道

（6）我们大多数人都应当改变有关个人性质和历时的个人同一性的观点。此处的真理与我们绝大多数人所相信的大为不同。

（7）鉴于我们有关我们自己的那些信念中的这一变化，我们应当改变我们的某些道德观点。而且某些极端主张，尽管它们能够遭到可以得到辩护的否定，但却是可以辩护的。

这些结论也难以评估。如果我们采纳那些极端主张，这些结论可能是不受欢迎的。斯温伯恩写道，如果他接受还原论的观点的话，他就不会有什么理由继续活下去了。其他一些学者主张，绝大多数道德则会从根本上遭到削弱。

我拒斥这些极端主张。我认为关系R——连续性和联系性——赋予我们特别关心我们自己的未来的一个理由。这个理由可能并不如那个进一步的事实所会提供的理由强。而且，因为心理的联系性是一个程度问题，所以我们应当拒斥这样一个主张：对我们的未来的某些部分更少一些关心必定是非理性的。我们应当拒斥古典自利论。我已经解释过我为什么欢迎这个结论。如果我们变成还原论者，那么这个观点改变也支持我们的那些道德观点方面的某些改变。但是我并未发觉这些令人担忧。

当我考虑结论（6）的蕴含的时候，（6）为真使我欣然不已。这种观点改变也有心理后果。它使我对我自己的未来，以及对我将有一死这一事实关心得少了一些。相比之下，我现在更加关心他人的生活了。我欢迎所产生的这些后果。形而上学会产生哲学的慰藉。

最后我论证道

（8）因为我们能轻易地影响到未来人的同一性，所以我们面对着那个非同一性问题。要想化解这个难题，我们需要一个有关善行的新理论。这个理论还必须避免那个令人讨厌的结论和那个荒谬的结论，而且要解决那个纯粹加法悖论。

既然我们尚未找到这个理论，那么这些结论是不受欢迎的。它们从根本上削弱了我们有关对未来的世世代代所负有的义务的信念。我们绝大多数人认为，对两个社会政策中的一个政策的选择或许有悖于那些生活在更遥远的未来的人们的利益。在许多情况下，这个信念是错误的。对我们的冒险政策选择或者对我们的消耗选择一定存在道德

诘难。但是，这个诘难不能诉诸我们有关伤及他人为错的平常原则。尽管这两个政策都具有明显的糟糕后果，但是选择这两个政策却不会对什么人来说将更糟。

既然我未能发现我们应当诉诸的那个理论，那么我就不能够解释对于我们选择这些政策的那个诘难。我认为，尽管我到目前为止失败了，我或者他人总会发现我们所需要的这个理论。但是在做到这一点之前，结论（8）是一个令人担忧的结论。

同时，我们应当向那些将要决定是否增加核能源使用的人们隐藏这个难题。这些人知道那个冒险政策可能造成更遥远的未来中的一些灾难。如果这些人错误地认为选择那项冒险政策会有悖于那些被这样的灾难夺去性命的人们的利益，那么将更好一些。如果他们持有这样的错误信念，他们将更有可能达到确当的结论。

结论（8）在别的一些方面也是不受欢迎的。我们绝大多数人会认为，那个令人讨厌的结论和那个荒谬的结论只不过是我如此称呼而已。如果我们不知道如何避免这两个结论，不知道如何化解那个非同一性问题和那个纯粹加法悖论，我们就将具有这样一些信念：对之我们不能提供正当性证明，而且知道它们是不一致的。

如果我或者他人解决这些问题，结论（8）将在微不足道的意义上受人欢迎。我们喜欢享受解决问题之快乐。但是在我们发现解决方法之前，我们应当对这个结论表示痛惜。伴随着更多的未解问题，我们进一步远离统一理论。我们进一步远离了化解我们的那些分歧的理论，而且鉴于它达成了这个目标，或许配得上称作真理。

153.道德怀疑论

道德怀疑论者否认道德理论可以为真。更宽泛些说，他们否认任何理论客观上会是最好的理论。辩护这个观点的一个论证说，伦理学不像数学，不是我们都认同的一个学科。我们能够否定这个论证是一个好论证。但是，要从根本上削弱这个论证，我们必须找到一个化解

我们的分歧的理论。在我们寻获这样的一个理论之前，我们能够给出别的两个对道德怀疑论加以置疑的根据。这些根据是用以主张客观性问题不是已经解决而是仍然悬而未决的。

许多人是道德怀疑论者，但并不是有关合理性的怀疑论者。如果我们不是仅仅考虑行动的道德理由而是考虑行动的所有种类的理由，那个客观性问题就能够得到最好的求索。有些主张是我们人人都接受的。

假设如果我不移动一下的话，我将会被落石砸死，而且我最想要的是活下去。我有移动的理由吗？无可否认，我们有。这个主张在所有时代、在所有的文明中都会被接受。这个主张是真的。

既然有一些有关行动理由的真主张，我们就能够否定一些怀疑论者所主张的。有时有人主张说，与岩石或者星星不同，不能够有客观的道德价值。这样的一些实体不能够存在。它们过于奇异，不可能是“宇宙构造”的一部分。但是在刚刚描述的那个事例中，我的确具有移动的理由。这可能不是一个道德理由。但是，既然有这么一个理由，就可能有理由。行动的理由在惟一相关的意义上能够“存在”。既然有一些行动理由，其中的一些是否是道德理由的问题就是一个有待讨论的问题。[2]

还有另一个置疑道德怀疑论的根据。我们不应当假定伦理学的客观性必须要么全有要么全无。可能有一部分道德是客观的。在描述这个部分方面，我们的主张可能是真的。当我们考虑这部分道德的时候，或者这些道德问题的时候，我们可能发现会消除我们的那些分歧的统一理论。可能有别的一些问题我们永远不会达成一致。这些问题可能没有什么真的答案。既然客观性无需要么全有要么全无，道德怀疑论者可能在部分上是对的。这些问题可能是主观的。但是这并不需要给统一理论投下一层怀疑的阴影。[3]

154.人类的历史和伦理学的历史何以可能刚刚开始

一些人认为，伦理学不可能有进步，因为一切都已经被说过了。像罗尔斯和内格尔一样，[4]我所认为的正好相反。有过多少人把非宗教伦理学当作他们的毕生事业？在最近的过去之前，为数极少。在绝大多数文明中，绝大多数人已经相信了有一个上帝存在，或者有多个神存在。一个相当大的少数派事实上是些无神论者，无论他们佯装是什么。但是，在最近的过去之前，几乎没有什么无神论者把伦理学作为他们的毕生事业。就像孔夫子和一些古希腊古罗马人可能是一样，佛陀也可能是这些为数不多的人中的一位。在经过1000多年之后，在16和20世纪之间又多出了一些。休谟是一位把伦理学当作他毕生事业之一部分的无神论者。西季威克是另一位。在西季威克之后，有数位职业道德哲学家。但是他们大多并不治伦理学。他们所治的是元伦理学。他们并不追问结果是好还是坏，或者行动是对还是错。他们所问和所写的只是关乎道德语言的意义和那个客观性问题。非宗教伦理学得到许多人的系统研究，还只是20世纪60年代之后的事。与其他科学相比较，非宗教伦理学是最年轻也是最不发达的一门科学。

我认为，如果我们毁掉人类，就像我们现在能够做得到的，这个结果将比绝大多数人所想的糟糕很多。比较一下三个结果：

(1）和平。
(2）一场杀死世界现存人口99%的核战争。
(3）一场杀死100%人口的核战争。

结论（2）会比（1）更糟，而（3）会比（2）更糟。这两个差别哪一个更大？绝大多数人认为更大的差异是在（1）和（2）之间。我认为（2）和（3）之间的差异大得很多。

我的观点由非常不同的两组人所持有。两组人都会诉诸同一个事实。地球将至少再有10亿年可供人类居住。文明只是在几千年之前才开始的。如果我们不毁掉人类，这寥寥的几千年可能只是整个人类文

明历史的一个小片段。于是结论（2）与（3）之间的差异可能就是这个小片段和这个历史的所有其他部分之间的差异。如果我们把这个可能的历史比作1天，那么迄今所发生的只不过是1秒的一个小数而已。

那些持有我的观点的两组人中有一组是古典功利主义者。他们会主张，就像西季威克所主张的，毁灭人类会是迄今可以想象出的最大的犯罪。这个罪行的糟糕性在于对可能的幸福总量的大规模毁灭。

另一组人会表示认同，但却出于非常不同的一些理由。这些人认为在纯粹的幸福总量中价值微乎其微。对这些人而言，重要的是西季威克所称的"理想的诸善"——科学、艺术和道德进步，或者向一个世界范围的完全公正的共同体的一种连续不断的迈进。毁灭人类会妨碍这三种善的进一步成就。这会是极端糟糕的，因为最重要的是这三类善的最高成就，而且这些最高成就会出现在未来的一些世纪中。

在为一个世界范围的完全公正的共同体而奋斗的过程中，显然会有更高的成就。而且在所有科学和艺术中都会有更高的成就。但是最大的进步会在这些科学或艺术中目前最不发达的那个领域取得。我已经说过，这就是非宗教伦理学。信仰上帝或者多神妨碍了道德推理的自由发展。不信上帝，得到多数人的公开容许，只是最近的一个事件，尚未完成。因为这个事件过于新近，所以非宗教伦理学尚处在一个非常早期的阶段。我们尚不能预言我们是否将会像在数学中那样在非宗教伦理学中达成共识。既然我们无法知道伦理学将如何发展，那么抱有高远的希望并非是非理性的。

---

[1]威廉斯针对行为功利主义提出一个论证，宣称在这个理论的结构中有一个断裂点。参见Williams（1），第114页。

[2]我们能够通过诉诸Sidgwick（1）第37—38页上所提出的非道德理由来诘难怀疑论者。

[3]我沿袭Nagel（3），第97—126页和Nagel（4），第9章和第14章。

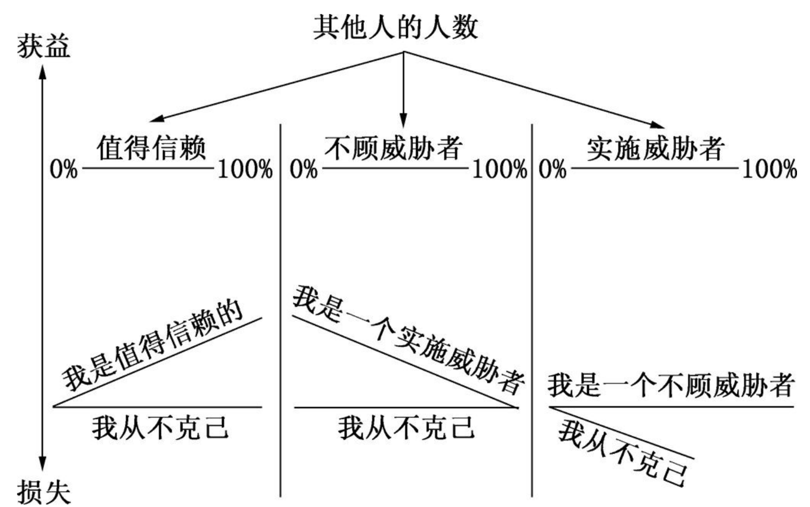
[4]我沿袭J·罗尔斯，《道德理论的独立性》，《美国哲学协会会报和发言汇编》（1974—1975年），第5—22页；以及Nagel（4），第9章。


附录
上海译文出版社

A 一个没有欺骗的世界

假设我们既坦率又从不克己。称之为现状。对我们每一个人来说，如果变成一个值得信赖的、实施威胁的不顾威胁者大概会更好一些。每一改变都会包含一些特定的风险。但是这些风险很有可能被那些利益大大地超过。

如果某个人在这三个方面改变他的意向的话，他会有什么样的获益？这取决于他人做什么。来自变得值得信赖这一点的获益取决于其他变得值得信赖的人的数目。来自变成一个实施威胁者这一点的获益取决于其他变成不顾威胁者的人的数目，反之亦然。如果别的每一个人都保持从不克己，那么做一个纯粹的不顾威胁者不会有什么裨益，而做一个值得信赖的人充其量也只会有一点好处，而做一个实施威胁者却会有很大的益处。如果没有别的什么人是值得信赖的，某个值得信赖的人则几乎没有什么获益，[1]如果别的每一个人都是值得信赖的，他则获益最多。如果没有别的什么人是实施威胁者，那么一个纯粹的不顾威胁者则毫无获益，而如果别的每一个人都是实施威胁者，他则获益最多。但是如果没有别的什么人是不顾威胁者，一个实施威胁者则获益最大。这些事实图解如下。



每一个人都被假定是坦率的，而且从不克己，除非他获得这三种意向中的任何一个。该图示忽略了那些风险和其他某些复杂之处。通过作出一些进一步的假定，我们会避免这些复杂之处。但是我们在这里无需如此，因为它们不会影响到论证。不顾威胁者是实施威胁者的一种特例，他发出了不顾他人的威胁的威胁。如果这是某个人惟一会兑现的威胁，我们则称这个人为纯粹的不顾威胁者。

就像图示所表明的，如果某个人要么从变得值得信赖而获益，要么从变成一个实施威胁者而获益，这些获益能够是与现状相关的获益。与倘若他和别的每一个人都保持永不克己的话相比，这样的一个人变得境况更好了一些。但是某人从成为一个纯粹的不顾威胁者而得到的获益并不能够把他提升到现状之上。它只能够防止他下降得更低。

这些事实能够得到如下解释。当某个人从变得值得信赖而获益的时候，他人也获益这一点时常也是真的。而且这些获益无需涉及对他人来说的损失。这些获益能够来自遵守互利的协议，这创造出不以他人的损失为代价的新的利益。例如，某些工业上或农业上的合作形式就是如此。但是，当某个人从成为一个实施威胁者而获益的时候，这对某个别的人来说则会更糟了。对该实施威胁者而言的获益可能只是一种防御性的获益，防止一个未来的侵略者从侵略中获益。但是这对于这个侵略者来说会更糟。而当某个人从成为一个不顾威胁者而获益的时候，这只是避免一种损失。如果我是一个坦率的不顾威胁者，那些实施威胁者则不能够通过威胁我而获益。如果他们威胁我的话，对他们而言情况将更糟，因为我将不顾他们的威胁，况且他们将实施那些威胁，这对我们所有的人而言全将更糟。既然如果他们威胁我的话，对他们而言更糟，他们将不会这么做。但如果我是一个纯粹的不顾威胁者，我惟一的获益就是这一类的获益——我并不输给实施威胁者们的获益。这就是为什么这个意向不能够把我抬高到现状之上的原因。

这些事实具有下述一些蕴含。如果每一个人变得值得信赖的话，那么对每一个人而言这会比倘若没有什么人这么做的话要更好。如果我们全都变成实施威胁的不顾威胁者的话，则没有这样的全面获益。这解释了为什么从这三种永不克己的意向出发，只有第一种情形是道德所要求的。如果我们能够影响我们的意向，我们则应该在道德上导致我们自己成为或者保持值得信赖，这是一个令人信服的主张。但是，如果现在我们在与威胁相关的方面全都永不克己，我们则应该在道德上导致我们自己成为实施威胁的不顾威胁者，这却不会是一个令人信服的主张。

尽管这个一般性的主张并不令人信服，但是我们可能做出另外两个主张。如果其他那些人具有足够坏的意图，我们大概应该成为坦率的实施威胁者，以便我们能够吓阻这些人。（如果我们不是坦率的，那么只是显得像实施威胁者在道德上会更好。就核威慑而言，这一点再清楚不过了。）我们还可以主张，如果其他那些人已经成了实施威胁者，而且有坏的企图，我们则应该在道德上导致我们自己成为坦率的不顾威胁者。

在现今之世界，我们是部分不坦率的，让人相信我们真的是兑现威胁的不顾威胁者会相当困难。以某种对我们自己来说的小代价来兑现或者无视某个威胁会是不够的。既然对我们来说显得是些实施威胁的不顾威胁者会更好，那么以期获得这个有用的外表，对我们来说以利己的标准衡量付出这种小的代价可能是合理的。但是正是这个事实会使人们怀疑我们会以对我们自己来说的巨大代价来实施或者不顾威胁。一个实施威胁者因而应当欢迎发展绝对可靠的测谎测试。

我们意向中的这三种变化所包含的风险是什么？如果我们不是坦率的，成为值得信赖的人的一个风险是，我们可能被那些只是显得值得信赖却并不尽其本分的人们骗得去遵守某个双边协议。如果我们全都是坦率的，那么就只有一个小得多的风险，那就是，那些我们与之达成双边协议的人们尽管打算尽其本分，事实上却可能无法做到。这

个风险会被那些很有可能的值得信赖的利益所大大地超过，那些利益是由遵守互利的协议所创造出来的。成为一个实施威胁者的风险是，我们可能威胁的是一个不顾威胁者；而成为一个不顾威胁者的风险是，我们可能受到一个实施威胁者的威胁。但是如果我们全都是坦率的，这两个风险会是小的。如果一个实施威胁者威胁的是一个坦率的不顾威胁者的话，对实施威胁者而言显然会更糟。而且我在假定的是，除了当我们是在按照这三种意向行动的时候之外，我们当中没有什么人会做那些他认为会对他自己更糟的事。风险只是这个假定可能偶尔未能遵守。

如果我们全都变得坦率地值得信赖，这对每一个人来说都会更好。我们可能会上升到现状之上。如果我们全都变成坦率的实施威胁者，这不会对每一个人来说都更好。这可能对某些人来说更好。这些是那些自然软弱的人。如果我们全都从不克己，那么强者能够不用威胁而用警告来压榨弱者。如果因为弱者没有作出让步而使强者伤害到弱者的话，这对强者而言可能并非更糟。但是如果弱者拥有本来可以一起毁灭掉他们自己和强者的武器的话，对他们而言成为坦率的实施威胁者可能更好。凭借发出可信的防御性的威胁，他们本可以使他们自己摆脱强者的压迫。对弱者来说的这一获益可能就是加卢瓦将军为什么欢迎核扩散的部分理由。但是这一获益会不保险。如果强者变成坦率的实施威胁者的话，它就会化为乌有。而如果我们全都变成兑现威胁的不顾威胁者的话，对每一个人来说风险就会更大。这会使我们意向中的这些变化对我们所有的人来说更糟。

最后这个主张可能看起来与我的另一些主张相冲突。我主张说过，如果我们每一个人导致他自己成为一个实施威胁的不顾威胁者的话，对我们每一个人而言会更好。如果这些变化对我们所有的人来说会更糟的话，那么可能看起来这不会是真的。

并非如此。下面这些主张两者都会是真的：（1）无论别人做什么，对我们每个人来说如果自己成为一个实施威胁的不顾威胁者的话

会更好；（2）如果我们全都而不是没有什么人作出这些改变的话，这对我们大家全都会更糟。通过作出这些改变每人之所获可能少于如果别的每一个人作出相同的改变的话他所会失去的。

假设我们处于现状，全都坦率地永不克己。如果我们添加一个假定，那么倘若他使自己成为一个实施威胁的不顾威胁者的话，对任何一个人来说不会更好。这个假定是这样的，就是说，如果任何一个人在这两个方面改变了自己的话，他会从而导致别的每一个人作出相同的两个改变。让我们假设这不会为真。如果我们的数目非常庞大——是一个人口庞大的社会中的成员，它就未必。

如果我不会被别的每个人所效法的话，倘若我变成一个坦率的实施威胁的不顾威胁者对我而言则将更好。如果别的每个人都保持永不克己，那么我成为一个不顾威胁者则毫无所获，但是我却从成为一个实施威胁者中所获最多。这就是我如何会升到现状之上的最高点的。随着其他人开始获得这两个本性，我从成为一个实施威胁者身上所获越来越少，但是我开始从成为一个不顾威胁者身上获益。当别的每一个人已经作出这两个改变的时候，我就会停止处于现状之上。由于存在某些风险，我甚至可能下沉得更低。而且我现在从成为一个实施威胁者身上毫无所获。既然现在别的每一个人在本性上都倾向于不顾我的那些威胁，那么它们就已经变得不管用了。而且我或许愚蠢地发出一个威胁的风险既然存在，那么倘若我失去这个本性可能现在对我而言更好。但是我现在从成为一个不顾威胁者身上所获最大。在一个充满实施威胁者的世界中永不克己对我来说会是非常糟糕的。那么别的每一个人都会通过发出威胁来压迫我。也许绝大多数人生性善良，不会这么做。但有些人会这么做。所以对我来说仍然做一个坦率的不顾威胁者显然会更好。

那么我的结论是这样的。在一个没有欺骗的世界中，如果我们各自在至少两个方面停止永不克己的话，对我们每个人而言非常可能地是会更好。如果各自成为并保持做一个值得信赖的不顾威胁者的话，

对各自而言非常可能的是会更好。根据自利论，对我们各自而言以这两种方式改变自己会是合理的。

正如我已经说过的，我们不能够使我们自己以那些我们认为是非理性的方式行动，这或许为真。而且别的人不会认为我们会以那些我们认为是非理性方式行动，这也或许为真。如果无论这两者中的哪一个为真的话，那么自利论会告诉我们，对我们来说不仅改变我们的意向而且改变我们有关合理性的信念会是理性的。我们各自应当仍然认为，去做那他认为对自己将更糟的事情通常是非理性的。但是他应当努力使他自己相信，当这样的行动牵扯到不顾威胁或者遵守诺言的时候，它们是合理的。

（至于我们或许经由技术途径而变得坦率，以及我们应当事先考虑这样的变化，我都是得自格洛弗。）要想更加深刻地思考这些说法，请参阅Glover（3）。

---

[1]B·胡克纠正了我先前的这样一个观点，那就是，如果某人是值得信赖的，那么倘若再没有别人是值得信赖的，他会一无所获。并不值得信赖的某人或许带给我们利益，因为他相信我会给予他某种利益作为回报。

B 我较弱的结论在实践中如何会击败自利论

按照我的两个结论中较弱的那一个结论，当S和P相冲突的时候，遵循其中的任何一个都是合理的。这个结论对自利论会有很高的毁坏力。从形式上说，该结果会是对称性的。自利论就会失掉其为那一个真的——或者最好的——理论的主张。它就会失掉那个更胆大的主张

（S12） 在明知违背一个人自己的自我利益的情况下，只是为了达到行动之时和理想的熟虑之后一个人最想要或最珍视的东西而行动是非理性的。

它会剩下

（S13） 在明知行动符合一个人自己的自我利益的情况下而行动是合理的，即使是在一个人明白这行动将阻碍在理想的熟虑之后其最想要或最珍视的事的时候。

当前目标论不会赢得那个更大胆的主张

（P11） 在一个人明白这行动将阻碍到在理想的熟虑之后其最想要或最珍视的事的时候而采取符合自己利益的行动是非理性的。

它只会赢得

（P12） 一个人去做明知将最好地达到其在理想的熟虑之后最想要或最珍视的事是合理的，即使是在一个人明知这违背自己的自我利益的时候。

尽管存在形式上的对称性，但是这些可比的主张对这两个理论的重要性却不同。对自利论而言做出它的主张中的那个更大胆的主张更为重要。丧失了（S13），它就丧失了它在与P的冲突中所需要的东西。它不足以保住那个较弱的（S12）。

在我辩护这个主张之前，我将先考虑一个有关自利论与道德之间冲突的相似的主张。西季威克认为，当我们问到我们最有理由做的是什么的时候，其结果在形式上是对称的。最好的道德理论不能够击败或被有关合理性的自利论所击败。当道德与自我利益相冲突的时候，遵循其中的任何一个都是合理的。[1]

如果是这样，那么这个结果对于双方的哪一方更具有毁坏性呢？许多人假定它对道德论的毁坏性会更大。是故休谟问道："什么道德理论能够达到有用的目的，除非它能够表明……它所推崇的那些义务还是各个个人的真正利益吗？"[2]休谟在此所蕴含的是，如果我们被迫选择的话，我们会选择自我利益。正如他所写的，"我们对于使人类从事一种我们承认充满着艰辛的实践能够抱有什么希望？"如果道德和自我利益会出现冲突的话——他蕴含着——这就会是我们的任务。休谟的正式结论是，这两者总是相契合的。[3]通过表明这一点，我们所表明的是道德论"所有的那些真正的、最迷人的魅力……那件令人忧郁的盛装跌落了"。

休谟的蕴含是虚假的。当道德与自我利益相冲突的时候我们全都选择后者这一点并不为真。有无数的例子，其中某人做他所认为道德上是对的事，即便出于他不信奉宗教之故，他所认为的是他是在做对他而言将更糟的事。纵使我们相信，在那些冲突的情形中，以两者中的任何一种方式行动都会是合理的，这个信念对道德论也并非明显地更具有毁坏性。

西季威克明白这一点。在陈述完这个信念之后，他写道："实践理性由于自身之中产生了矛盾，它可能不再成其为这一方面或另一方面的一个动机。该冲突也将不得不由这两组非理性冲动中较优势的一组来决定。"[4]

西季威克并未讨论当前目标论。如果这个理论进入该领域，而且结果是三方平局，那么对西季威克的问题则有一个自然的回答。如果

三个竞争对手中没有哪一个比另外两个提供更强的理由，当三个之中的两个之间相一致时，它们或许比第三个提供更强的理由。如果是这样，我们就能够回答我们最有理由做的是什么这个问题。假设我思虑清楚，而且知道，如果我尽我的义务对我将更糟。如果我最想要的是尽我的义务，那么这是我最有理由去做的事。如果代之我最想要的是对我将更好的事，那么这是我最有理由去做的。我们必须承认，我不会有什么进一步的理由来据以拥有这些愿望中的一个而不是另一个。这得自这样一个假定，就是当道德与自我利益相冲突的时候，其中没有哪一个提供更强一些的行动理由。尽管我对一方的偏爱不会是合理地所要求的，但是它不会是非理性的。正如西季威克所写到的，它会是"非理性的"。而我的偏爱是非理性的这一点并不使之成为随意性的。它并不会像从完全相似的两个事物之中随机地择取其一那样。西季威克认为，当道德与自我利益相冲突的时候，我们不能具有更多的理由来遵循其中的任何一个。我则业已主张我们能够。

现在回到那个相似的结论，就是当P与S相冲突的时候，其中没有哪一个击败另一个。按照这个结论，当这些理论冲突的时候，遵循其中的任何一个都是合理的。

我在此所主张的是我因休谟有关S与道德相冲突的主张对他所做的批评。如果结果是平局，那么这对这两个理论中的一个更具有毁坏性。它对S的毁坏性更大。这是因为，在这两个理论之中，S需要更具批判性。我们自然倾向于做我们知道将最好地实现我们的当前目标的事。但我们并不肯定这么做。其中我们未能做到我们按照当前目标论最有理由做的事的那些方式能够有许多。但是如果我们并不认为我们是在非理性地行动，那么在这两个理论中我们将更有可能予以遵循的则正是这个理论。

我之所以可能以某个方式行动，是因为我认为不这么做会是非理性的。用西季威克的话来说，"实践理性"本身能够是"一个动机"。只

是因为我们具有一个不想非理性地行动的愿望，我们才无需裁定这是否为真。显然，我可能具有这个愿望，而且这会使我有关合理性的信念影响到我的行动。假设我要么能够（1）做将最好地实现我在知道事实且思虑清楚的时刻最想要或者最珍视的东西的事，要么能够（2）做将符合我自己的长远利益的事。而且假设我认为，当S和P相冲突的时候，没有哪一个获胜。我认为，在这样的情形中，无论做（1）还是（2）都会是合理的。如果我认为其中没有哪一个行动会是非理性的，那么我将更有可能做什么则是显而易见的。我将更有可能做会最好地实现在行动的时刻我最想要或者最珍视的事。那么我所在遵循的将不是S而是P。

假设相反我认为当这两个理论相冲突的时候，S击败P。代之以（S13），我接受更大胆的（S12）。那么我可能做（2）而不是（1）。我具有一个不想非理性地行动的愿望，而且，既然（1）对我而言会更糟，那么我现在认为：（1）会是非理性的。如此一来，我的有关合理性的信念可能导致我遵循S。

可能有这样的诘难，就是，如果我现在之所以做（2）是因为我认为只有（2）是合理的，那么我必定仍然是在按照我的最强烈的愿望在行动，或者是在做最好地实现我的当前愿望的事。这个主张具有争议性。如果它为真，那么这两个理论在此就会相契合。但只是因为我认为自利论而不是当前目标论，这才会为真。如果我代之认为当前目标论，或者认为其中没有哪一个理论击败另一个，这两个理论就会相冲突。

当这两个理论相冲突的时候，我们将自然倾向于做那将最好地达到我们的那些当前目标的事。这就是为什么在这两个理论之中自利论需要做出那个更大胆的主张的原因。它需要这样一个主张，那就是，当我知道我去做那会最好地达到我的那些当前目标的行动对我而言会更糟的时候，我去做这个行动则会是非理性的。如果我认为这会是非理性的，那么我可能被引导到去做那对我会更好的事。但是，如果我

认为其中没有哪一个行动是非理性的，那么我将自然倾向于做那将最好地达到我的那些当前目标的事。而且这适用于我们所有的人。当我们合理地行动的愿望或者避免非理性的愿望并未告诉走两者中的哪一条路的时候，我们将更有可能努力去做那将最好地实现其时我们最想要或者最珍视的东西的事。这就是为什么说在当前目标论与自利论的冲突中，理论上的平局对当前目标论而言实际上是一种胜利的原因。

---

[1]G·哈曼置疑这是否是一个可以得到辩护的观点。有人可能说："道德和自我利益两者都提供行动理由。这些理由在不同的情形中具有不同的力度。当道德与自我利益相冲突的时候，对于什么是合理的这个问题可能有答案。一个非常强的利己的理由在权重上超过一个弱的道德理由，而且反之亦然。只有在某些情形中才没有哪种理由比另一种理由更强一些。"按照这个观点，这两种理由总能够按某种中立的尺度在权重上被超过。而答案在不同的情况下却不同。存在以不同权重度量的公度性。

假设我们采纳常识道德的某个版本。那么我们会回应说："这些主张是不对的。任何一个令人信服的道德理论都把行动者自己的自我利益考虑进去。假设我答应给你某个相当微不足道的利益。那么这样一点就变得是真的，就是说，要遵守这个诺言我必须遭受巨大的损失。这不是一个弱的道德理由被强的利己理由在权重上所超过的事例。在这个事例中，当我只有付出这个非常大的代价才能遵守这个诺言这一点变得为真的时候，我应当遵守这个诺言这一点就不再为真。在这样的一个事例中，道德并不与自我利益相冲突。相似的主张适用于所有道德原则。我们是否应当按照这些原则行动部分地取决于我们得要做出多么大的牺牲。当要求我们做出太大的牺牲或者过于影响到我们的生活的时候，那些不然会成为我们的道德义务的东西便不再是我们的义务。既然道德给予那些自我利益主张以应得的权重，那么，你就不能够主张说一个弱的道德理由会被一个强的利己理由在权重上所超过。倘若即便会付出巨大的代价你也应当以某个特定的方式行动，那么你所要付出的代价就已经被赋予它应得的权重。尽管有这个代价，但这是你应该去做的。"按照这个观点，那些利己的理由是由道德尺度来权衡的。而且不存在一个能够对这两种理由加以权衡的中立尺度。在那些接受这个观点的人中有一些拒斥这种自利论。他们认为，当道德与自我利益相冲突的时候，追求自我利益是非理性的。但另有一些采纳西季威克的观点。这些人们认为，当这两者相冲突的时候，无论遵循哪一个都是合理的。

下面假设我们接受某种形式的后果论。按照这个道德理论，行动理由是行动者中性的。不能够主张说，某个人是否应当以某个特定的方式行动取决于他特别（he in particular）要牺牲的有多大。行动者的利益没有什么特别的权重。他不享有什么大于任何一个别人之利益的权重。后果论者可能主张，按照他们的理论，那些自我利益主张已经被赋予应得的权重。但是不能在常识道德能够作此主张的那个意义上作此主张。在决定行动者应该做什么的时候，常识道德确实赋予行动者自己的自我利益以特殊的权重。它因而可能主张说，它赋予利己理由的特殊力量以应得的权重。后果论者则不能够做出这一主张。

后果论者尽管不能够做出这个主张，但是他们可能也拒斥有关存在着以不同权重度量的公度性这个观点。他们能够指出，既然C是行动者中性的，它提供了一种完全不同于利己理由的一些理由。他们可能主张说，因为这些理由是如此地不同，既然没有什么中立的尺度，以至于它们不能够彼此分出轻重。

[2]Hume（3），结论，第2节。

[3]在西季威克之前，这几乎总是人们所提出的解决之道。休谟的新颖之处是不经诉诸来生而提出这一点。我之所以称休谟的解决方法是"正式的"，是因为他的话可能是具有讽刺意味的。他写道：

即便最直率地对待罪恶，而且做出所有可能的让步，我们也必须承认，在任何情况下都没有哪怕最小的借口、出于自我利益的观点赋予罪恶对于美德之优先地位；除非大概在正义的情况下——其中一个以一种特定的眼光对待事物的人可能时常看起来就他的诚实而言是一个失败者。尽管不涉及财产就没有什么社会可以存在这一点得到承认，但是按照人类事务所进行的那种不完美的方式，一个明智的小人在特定的情况下可能认为不公或者不实的一个行动将为他添加相当可观的好运……诚实为上可能是一个良好的一般性原则，但是面临许多例外；而他，人们也可能这么认为，最有智慧地指导他自己的行动，既遵守了那个一般性的规则，又利用了所有的例外。

我必须承认，如果一个人认为这个推理非常需要一个答案，那么找到任何一个对于他足够充分的答案便不会有多大困难。

"对于他"这些词得到正当证明吗？当记起休谟先前的一些话的时候，我们可能对此有所怀疑。休谟说过："只怀有不正心术的人无论提出的理论有多真，他必须承认，它所导致的是一种危险的和有害的实践……为什么再从埋葬瘟疫的坑中把它挖掘出来呢？那些对于社会而言有害的真理，即便有这样的真理存在，将产生一些错误，后者是有益和有利的。"这些话提醒我们，休谟后来的主张可能不是真的。正如西季威克描述他年轻的时候所写到的："如果我对个人的幸福和一般的幸福之间的契合性发生怀疑，我倾向于认为应当果决地把这种怀疑弃之不顾。"（Sidgwick（1），第xv页）但是西季威克后来既未能欺骗他自己也未能骗过公众。看起来休谟也不大可能欺骗他自己。

在承认他没有什么推理可以回答那个明智的小人之后，休谟写道："如果他的内心并不反抗这样的有害准则，如果他对邪恶和卑劣的想法没有感到什么迟疑……"道德凭其本身恐怕不能击败自利论。它恐怕必须诉诸人心，并且与当前目标论结盟。如果这些盟友击败自利论，它们的结盟就将终结。当前目标论与道德时有冲突。但是，如果它是道德之对手，那么对于道德的威胁可能要少一些。对于这一点我希望另择它处加以论证。

[4]Sidgwick（1），第508页。

C 合理性与有关自我利益的不同理论

我主张（1）较之于那个对自身的偏向，存在数个别的愿望或者关心模式，它们绝不较少合理性。而且我得出结论（2），如果某人具有这些别的愿望之一，那么他据此行动并不会较少合理性，即使他知道这将有悖于他自己的自我利益。

如果我们接受（1），那么我们必定最终接受（2）。但是我现在将描述我们如何能够通过改变我们有关自我利益的观点而延缓接受（2）。我们能够延迟我们从S前移到P。

我将描述一个发端于边沁之观点的思想路线。如果我们接受某个别的观点，我们则将在一个较晚的阶段切入这个思想路线。边沁的观点把S与关于自我利益的快乐论结合了起来。按照这个观点，各人最有理由做的事情是那将使他尽可能地幸福的事情。

这个观点时常奠基于心理自我主义的那个快乐论的版本，即这样一个主张，就是各人最想要或者在冷静的时候会最想要的是尽可能地幸福。这个主张是虚假的。我们经常有别的愿望，这种愿望即便在经过我们冷静的深思熟虑之后仍然会是我们最强烈的愿望。

实现或者追求这样的一些别的愿望可能是我们主要的幸福源泉。但是既然这些愿望所期待的并非我们自己的幸福，而是某种别的东西，按照这些愿望行动有时候将使我们较少幸福。当我们将永远不知道这些愿望是否得到实现的时候，这最有可能为真；但是，即便当我们将会知道的时候，这也时常为真。

按照边沁的观点，在这样的一些情形下按照这些愿望去行动是非理性的。只有当我们认为这些愿望是非理性的时候我们才将接受这个主张。一些人相信这一点。但是就许多这样的愿望而言，我们可能与之有分歧。

就那些对他人幸福的愿望而言，我们最有可能出现与之不同的看法。如果惟有这些愿望是别的我们所相信的并非非理性的愿望，那么我们不在有关自我利益问题上而在有关合理性问题上会与边沁不一致。因为这些愿望的对象仍然是某人的幸福，我们会继续接受有关自我利益的快乐论。但是我们会对S加以限定。我们会主张，当我们牺牲自己的幸福从而能够带给他人更大的幸福的时候，那么牺牲掉自己的幸福并非是非理性的。这是西季威克的观点。

其次假设我们认为有范围更广得多的一组愿望并非是非理性的。这些愿望的对象绝不是哪一个人的幸福。其中的一大类或许是些仰赖于特定的价值判断或理想的愿望。但是第60节表明，以其对待这些愿望来看，S必须被拒斥。另一大类我称之为成就欲（desires for achievement）。这些是成功地做到我们在工作中或在更积极的休闲中所努力做的事的一些愿望。是故，艺术家、园丁、木匠抑或任何种类的创造者都可能强烈地想要他的创作尽可能地好。他最强烈的愿望可能是创作一件杰作，无论是用油彩、花卉、木头还是辞藻。而一位科学家或者哲学家则可能强烈地想要做出某个重大发现或者推进人类智慧的进步。

这样的一些愿望可能是自立的，不是达到别的一些愿望的手段。如果我们实现这些愿望，固然可能时常促进我们自己和他人的快乐。但这并不是这些愿望的对象。况且还有许多情形，诸如追求特定种类的知识，其中这些愿望的实现在促使他人幸福方面将毫无作用。（用流行的讲法，我们还可能不仅想要实现这些愿望，而且想要使我们的成就得到承认。我们可能想要名望。但这是一个单独的愿望，而且可能更弱一些。而我们想要得到承认，可能只是因为如果没有它我们则不能够肯定自己是否真的已经创作出一件杰作，或者作出了一个重大发现。）

在努力实现这些追求成功的愿望方面，我们有时候将使自己较少幸福。即使我们知道这些愿望正在得到实现，这也可能是真的。是故

乔治·艾略特尽管明白自己是一个成功的小说家，但是她仍然总是不满足于已经取得的成就。[1]为成就而奋斗可能并不是那类本身即回报的奋斗，而可能是一场磨难。

假设我们断定这些取得成就的愿望中有些并不是非理性的。我们得出的结论是，按照这些愿望行动不是非理性的，即便我们知道这将使我们较少幸福，而且无法赋予其他人更大幸福。（就一个成功的小说家而言这可能不会为真，但就一个成功的科学家或者哲学家而言则可能为真。）既然我们已经裁定这些行动不是非理性的，那么我们现在不仅有根据拒斥边沁的观点，而且有根据拒斥西季威克的观点。我们有根据拒斥任何一个本身是或者包含有S的那个快乐论的版本的观点。[2]

有两种可能性。我们本可以改变我们有关合理性的观点，从S转到P。可替代的选择则是改变我们关于自我利益的观点。我们本可以从快乐论转到愿望实现论。

按照愿望实现论，去做使我们较少幸福的事可能并不违背我们的利益。无论是否使我们幸福，实现一个强烈愿望现在都直接算作符合我们的利益。在奋力解决他的难题的过程中，一个科学家或者哲学家，除了一些罕见的情形之外，会是悲惨的。而且这样一点可能是真的，那是，如果他较少努力一些的话，他本会较少一些悲惨。在此他的幸福可能真的有所牺牲。但是，如果他是在实现他最强烈的愿望，那么我们现在能够主张说，他的行动并不是在违背他的利益。我们裁定了按照这样的追求成就的愿望行动并不是非理性的，即便是处在这使我们较少幸福的时候。既然我们已经改变了有关自我利益的观点，那么这个裁定不再赋予我们据以对S加以拒斥的根据。

正如这个例子所表明的，如果我们从快乐论转到愿望实现论，我们就更无需从S转到P。这是因为，鉴于这一转变，S与P更加经常契合。但是它们仍然是不同的理论。本质上的差异再一次是关乎时间。如果我们假定愿望实现论，S则主张各人最有理由做的事是那将最好

地实现或者能够最好地使他实现其一生中的所有愿望的事。他未来的愿望就像他当前的愿望一样重要来计算。但是P只是诉诸一个人的当前愿望——诉诸他在理想的熟思之后行动时要么去做、要么想要的东西。

S和P时常相契合。在很多情况下，那会最好地实现某人的当前愿望的东西不会与他的未来愿望的实现相冲突。而且可能的确是，某人最强烈的愿望终其一生都是相同的。尽管S和P时常相契合，但是也有它们彼此冲突的许多情况。

一些这种情况涉及到那些较不关心他们更遥远的未来的人们。但是出于论证的目的考虑，我现在假设的是，我们已经谴责对于近期之偏向。我们是在考虑P的这样一个版本，这个版本主张，一个合理的行动者在对自己的关心中应当同等关心他的未来的所有一切。即便P提出了这个主张，它还是时常与S相冲突。这之所以为真，主要是因为就许多人的情形而言，他们的那些最强烈的愿望中的一些并不持续终生。许多人的最强烈的愿望是长时间持续的。但是很少有人的最强烈愿望总是始终如一的。这样我的成就欲可能始终如一，但是在不同的时候我可能爱着不同的人，可能支持不同的政治运动。抑或我可能总是爱着相同的一些人而且支持相同的政治运动，但是在不同的时间具有不同的成就欲。无论是哪一种情况，P都可能与S相冲突。那会最好地实现我的当前愿望的事可能并不与那会最好地实现或者能够使我最好地实现我终生的所有愿望的事相契合。

在这样的一些情况下，S主张，对我而言去做会最好地实现我的当前愿望的事会是非理性的。但是，如果这些是对于某些特定的成就的愿望，那么我们裁定它们并不是非理性的，而且依据它们行动并非是非理性的。我们因而具有一些新的根据，据以拒斥S。

如前所述，我们有两个选择。我们可以改变我们有关合理性的观点，从S转到P。或者我们再次改变我们有关自我利益的观点。我们可以从愿望实现论转到目标列表理论。

根据这个理论，无论我们的愿望是什么，对我们而言某些事物要么是好的要么是坏的。那些好的事物之一可能是某些种类的成就欲。根据目标列表理论，如果我实现我的成就欲，对我而言可能更好，即便这造成我的那些终生的愿望总体上来说实现得要差一些。我们裁定，以这种方式行动不是非理性的。既然我们已经改变了我们有关自我利益的观点，这个裁定就不再给予我们据以对S加以拒斥的根据。

当我们从快乐论转到愿望实现论的时候，这降低了但并未消除S与P之间的冲突。就从愿望实现论转到目标列表理论而言我们则不能够做出这样明确的主张。如果我们作出这个转变，这将改变那些其中S与P相冲突的情况。我们刚刚给出的情形就是一个例子。但是没有什么明显的根据来主张说，如果我们作出这个转变，S与P相冲突的情形就将更少一些。

S多么经常地与P冲突取决于目标列表的内容。此处的重要问题在于，按照目标列表理论，S与P是否总会相契合。我认为，按照该理论的任何一个令人信服的版本，这不会是真的。按照目标列表理论，某些种类的成就可能是那些对我们而言是好的而且使我们的生活过得更好的东西之一。但是对我们而言是好的那些东西还有别的一些，诸如两个成年人之间双向的爱、生养并爱子女、发展整体能力和意识到所有种类的美等。这样的一点可能是真的，就是为了实现我的成就欲，我必须放弃其他大多数对我而言是好的东西。如此一来，我可能是在做照客观清单理论而言总体上说来将对我更糟的事情。但是，既然我是在实现这个成就欲，我们裁定我不是在非理性地行动。我们因而具有对S加以拒斥的新的根据。

在这个思路的较早的两点上我们有过可供选择的余地。我们要么可以改变我们有关合理性的观点，要么可以改变有关自我利益的观点。在我们现在所已到达的这一点上，我们不再具有可供选择的余地。现在只有一个有待得出的结论。我们不能够在我们有关自我利益的观点方面作进一步的改变。我们必须改变我们有关合理性的观点。

我们必须拒斥S而采纳P。正如我所主张了的，如果我们认为别的愿望并不比那个对自身的偏向更少合乎理性，那么我们最终必须拒斥自利论。

当我们拒斥这个理论的时候，我们失去了改变我们有关自我利益的观点的动机。我们从而应当重新考虑这些改变。

我是以假设我们接受快乐论而开始的，认为较少幸福有悖于我们的利益。这里有一个与之不同的事例。透纳[3]想要他的画作在一个独立的美术馆中集中展出。假设我们裁定，对于透纳来说努力确保这个愿望得以实现并不是非理性的，即便要以损失他自己的某种幸福为代价。当我们采纳S的时候我们不能够辩护这个信念，除非我们从快乐论转到愿望实现论。那么我们就能够主张，透纳的愿望的实现符合他的利益，即便它使其较少幸福。这就允许我们主张说，在努力实现这个愿望方面，尽管是以他的某种幸福为代价，但是透纳并不是在非理性地行动。但是我们不得不得出这样的结论，就是在他谢世1个多世纪之后我们拒不把透纳的作品放到一个独立的美术馆之中，那么我们现在的行动是在违背透纳的利益，或者说我们是在做某种对他来说糟糕的事情。而且我们可能发现这个主张不令人信服。那么我们则拥有一些相互冲突的信念。

如果我们抛弃S，那么这种冲突就消失了。我们现在既能够主张说，纵使透纳在努力实现建立一个透纳美术馆方面有悖于他的利益的话，他也不会是非理性的，又能够主张说，我们现在拒绝建立这样的一个美术馆并不是在伤害透纳，不是在做某种对他来说糟糕的事情。

相似的推理更有力地适用于我们在愿望实现论的两个版本之间——那个非限定的版本与成功论之间——的选择。我们不能够令人信服地主张说，我的所有愿望的实现都符合我的利益。追忆一下在火车旅行之后我同情某个陌生人而且强烈地想要她成功的那个事例。做出这样的主张并不令人信服，即如果这个陌生人后来的确成功了，但是

我并不知道的话，这对我而言仍然是好的。做出这样的主张更加令人信服，那就是，对我而言更好的只是我自己的生活以我想要的方式或者我会想要的方式进行。如果我们诉诸愿望实现论的这个受限的版本——成功论——那么将有更多的事例，其中按照我的当前愿望行动则是在做出对我而言更糟的事。这些将是这样一些事例，其中我的愿望并非关于我自己的生活。如果我们认为这些愿望中的一些不是非理性的，那么我们将认为按照它们来行动并非是非理性的。如果我们接受S，那么我们就可能被引导到接受有关自我利益的非限定的愿望实现论。只有按照这个理论，这些愿望的实现才符合我的利益。但是如果这些愿望并不是关于我自己生活的，而且我永远不知道它们得以实现，那么，我们可能发觉难以相信这些愿望的实现对我而言将是好的。我们可能再次拥有相互冲突的信念。

如果我们从S转到P，冲突再次消失。我们现在能够主张，对我来说按照这些愿望行动并非是非理性的，即便这样做对我而言将是糟糕的。既然我们已经拒斥了S，我们已经失去作出这样一个主张的动机，那就是，如果这些愿望不是非理性的，它们的实现对我而言将是好的。我们已经失去接受有关自我利益的那个非限定的愿望实现论的动机。我们可以转到那个更加令人信服的成功论。[4]

---

[1]参见Haight通篇。

[2]可能遭到这样的诘难：“你认为（1）成就欲不是非理性的，而且（2）按照这样的一些愿望行动，即使对于我们自己的幸福造成一种已知的代价，也因而不是非理性的。而且你主张说这些信念是一些拒斥西季威克观点的根据。情况并非如此。倘若我们对事物有强烈的愿望，而不是对于我们自己的或者他人的幸福有强烈的愿望，我们将更幸福。如果我们具有很多成就欲，总体上说来我们将更幸福。即便这些愿望有时将导致我们自己较少幸福，这也是真的。鉴于这些事实，西季威克的观点蕴含着与你的信念（1）和（2）相一致的主张。西季威克的观点蕴含着（3）我们造成我们自己具有或者保持那些成就欲是合理的。而且它蕴含着（4），当我们付出自己幸福的一种已知代价来按照这些愿望行动的时候，这些行动并不表明我们是非理性的。这是因为我们是在按照这样一些愿望行动，对于我们而言造成这些愿望的丧失会是非理性的。”

B·胡克所提出的这个诘难诉诸我在第一章中所做出的主张。它减弱了我们拒斥西季威克观点的根据，但是它并未消除这些根据。倘若我们仍然记得我的另外两个结论，这一点就清楚

了。我论证过（5），我们造成自己拥有这些愿望是合理的这个事实并不表明我们的愿望是合理的。而且我论证过（6），因为我们是在按照这些愿望行动所以我们不是非理性的这个事实并不表明我们的行动是合理的。（5）和（6）在那个想象出的事例中得到例证，其中因为我无视你的威胁，你与我同归于尽。在这个事例中，我造成我自己具有无视你的威胁这么一个非理性的愿望是合理的。而且当我按照这个愿望行动的时候，即便可以声称我不是非理性的，但是我的行动是非理性的。

当我们记得（5）和（6）的时候，我们明白为什么仍然有根据拒斥西季威克的观点。这个观点蕴含着我们造成我们自己拥有或者保持成就欲是合理的。但是这并不意味着我们这样一个信念，即这些愿望是合理的。西季威克的观点还意味着，当我们以我们幸福的一个已知的代价来按照这些愿望行动的时候，我们不是非理性的。但是这并不意味着我们这样一个信念，即这些行动并不是非理性的。而且按照西季威克的观点，这些行动就像我无视你的威胁一样必定被宣称为是非理性的。既然我们认为这些行动并不是非理性的，我们就有根据拒斥西季威克的观点。

[3]透纳（Joseph Mallord William Turner，1775—1851），英国画家，擅长水彩画和油画。代表作有《运输船的遇难》、《雨、蒸汽和速度》和《迪埃普港》等。——译者

[4]现在产生一个有关行动者的那些愿望的道德重要性的问题，我们判定这些愿望不是非理性的，即便它们的实现并不符合他或者别的任何一个人的利益。我们应当赋予这样的愿望什么道德权重？如果我们相信我们应当赋予这些愿望的实现以某些权重，我们就不能够诉诸我们熟悉的善行原则。我们必须诉诸某个别的原则。我们可能代之主张，尽管某些种类的愿望为具有这个愿望的那个人提供好的行动理由，但是并不为其他人提供道德理由。这就是Nagel（3）第121—126中所提出的观点。

D 内格尔的大脑

内格尔相信他实质上是他的大脑。他为这个信念提供了3个论证。这些论证包含在一个未经发表的草稿中，他可能会予以修正；但是即便是在这样的一种草稿形式中，它们仍然值得注意。我将引述这个草稿。

内格尔的两个论证诉诸的一个有关意义、指称和必然性的观点。我们所指称的对象具有某些实质属性：这些对象必须具有的属性，因为如果没有这些属性它们就不会存在了。一些属性之所以是实质属性，是因为我们的语词的意义。如是，因为“三角形”的意义，三角形具有三个边就成了三角形的一个实质属性。但是按照我现在将要讨论的那个观点，对象还有一个不同的具有实质属性的方式。这些属性之所以是实质属性不是因为我们语词的意义。在我们发现有关我们正在指的东西的事实之时，我们发现这些实质属性。例如，按照这个观点，我们已经发现金子的一个实质属性是它所具有的原子数是79。每一具有这个原子数的物质必定是金子，而且没有什么不具有这个原子数的物质会是金子。这不是“金子”这个语词的意义的组成部分。[1]

当我们使用“人”这个词和“我”这个词的时候我们是在指什么？内格尔写道：“我所是者乃是任何事实上使托马斯·内格尔这个人与他自己和他的心理状态同一和再同一起来的东西。”我所是者乃是任何解释我的心理生活的心理连续性的东西。而且内格尔相似地主张说，这对解释是什么这一点悬而未决。如果连续性的载体是一种笛卡儿式的自我，那么那就是我真正之所是。他继续写道：“如果另一方面我的大脑的某些状态和活动是心理官能的基质，那么处于这些状态中的那个大脑……就是我之所是，而我的大脑毁灭之后我的存在问题是不可想象的。不过，我可能并不知道那是不可想象的，因为我可能并不知道我自己的同一性的那些条件。”正如他后来所写到的：“在努力构想在我

的大脑毁灭之后我的存在的时候，如果我事实上是我的大脑，那么在这样的情形中我将不能够成功地指称我自己。”

内格尔是一个还原论者。他同意个人的同一性并不涉及那个“进一步的事实”，这个事实在任何可以想象的情形中要么完全成立要么完全不成立。个人同一性只是涉及身体的和心理的连续性。尽管他是一个还原论者，但是内格尔在两个方面不同于我和其他一些人所辩护的那个观点。按照此一观点，根本重要的是关系R：心理的连续性和联系性。而按照内格尔的观点，重要的是个人同一性。而且因为他认为他就是他的大脑，他认为根本重要的东西是这个大脑的连续存在。

这看起来可能只是有关想象性的那些事例的一个分歧。没有同一大脑的连续存在，实际上就不存在什么心理连续性的情形。如果该分歧只是关于想象性的事例的话，那么就几乎不值得讨论了。但是这个分歧还涵盖实际情形，以及我们自己的生活。以我之见，重要的那两个关系之一，即心理的联系性，在时间中呈现递减程度。这一点是第十四章中我针对自利论所作的论证的一个实质前提。如果内格尔的观点为真的话，这个论证就会从根本上遭到削弱。在我们的实际生活中，同一大脑的连续存在不是一个程度问题。

这个主张需要一个限定。内格尔对一个重要的问题存而不论。回顾一下威廉斯的例子，其中那位外科医生为了消除所有的心理连续性而篡改我的大脑。作为结果产生的那个人会是我吗？尽管我的大脑遭到篡改，但显然还就是那同一个大脑。如果我是我的大脑，那么我将仍然存在。但是在上面所引述的一句话中，内格尔提到，我之所是并非我的大脑，而是处在某些状态中的我的大脑。这大概是提供心理连续性的那些状态。按照他的观点的这个版本，在威廉斯的例子的最后，我就不会存在了。

内格尔的观点的两个版本或许可以重新表述如下。按照那个更简单的版本，我之所是就是正常地造成我的心理连续性的东西。但即便当其并不造成心理连续性的时候我也会是这个东西。我们业已发现，

我之所是就是我的大脑。在威廉斯的例子的最后，那位外科医生已经消除了所有的心理连续性。但是既然我的大脑将仍然存在，我就将仍然存在。

按照内格尔观点的那个较不那么简单的版本，我之所是就是处于使其成为这一原因的那些特定状态中的、那种造成我的心理连续性的东西。根据该观点的这个版本，我的同一性版本不只是涉及到我的大脑的连续存在。它还涉及到心理的连续性。这个版本与上面讨论到的一个观点相契合：即狭义心理准则，这个准则所诉诸的是具有正常原因的心理连续性。

只有内格尔观点的第一个版本在实际情况中与我所辩护的那个观点不一致。我们应当接受该观点的这个版本吗？这部分取决于内格尔对"人"和"我"这些词语的意义的描述是否正确。还有另一复杂之处。在他和别人用"我"这个词所指的意义方面，内格尔所做的主张有两个。其一是他用"我"意指无论什么解释他的心理连续性的东西。其二是他用"我"意指他的经验的"没有观察到的主体"。

我从第二个主张开始。这难以否定。我不是思想、行动和经验的一个系列。我是我的思想的思想者，而且我是我的行动的行动者。我是我的所有经验的主体，或者说我是具有这些经验的那个人。

内格尔主张，当我使用"我"这个词意指我自己，亦即我的经验的主体的时候，我事实上是在指我的大脑。我们应当接受这个主张吗？

我们首先应当注意到，一个所打算的指称可能失败。称我们所力图指称的为我们所意图的所指。可能有某个对象符合那些我们有关我们所意图的所指的信念中的一个。但是这可能不足以使之成为我们所指称的对象。在关于我们所意图的所指方面，我们可能还有别的许许多多的信念，当它们应用于这个对象的时候则会是虚假的。那么我们就不会是在指称这个对象。而且我们可能是在什么也没有指称。[2]

一个例子会是这样的。古希腊人相信宙斯神是电闪雷鸣的原因。宙斯并不存在，而且希腊词"宙斯"并不指称任何东西。我们不应当主

张说，既然古希腊人相信宙斯是电闪雷鸣的原因，而这个原因是云的一种带电状态，那么宙斯就是这样的一种云，而且此乃希腊词“宙斯”所指称的。一种云太不像一个神，以至于不能成为希腊词“宙斯”一词的所指。

在使用“我”这个词的时候，我意图指称我自己，即我的经验的那个主体。内格尔认为，当使用“我”的时候，我们绝大多数人都具有一些有关我们所意图的所指的虚假信念。纵使我们并未意识到这一点，绝大多数人认为我们的同一性必定是决定了的。我们认为我们是些实体，其连续存在“要么全有要么全无”。如果我们各自是一个笛卡儿式的自我的话，这个信念本来会是真的。但是内格尔认为，没有什么这样的实体。没有什么实体带有我们认为我们的经验的那个主体所拥有的那些特殊的属性。在这个方面，情形就像希腊词“宙斯”的事例一样。但是内格尔主张，在使用“我”的时候，我们并非未能指称。“我”事实上所指称的是我的大脑。内格尔承认，我们的大脑并不具有我们认为“我”所指称的东西所具有的那些特殊属性。尽管一种云太不像古希腊人所认为的“宙斯”所指称的东西，但是内格尔主张说，我们的大脑则并非太不像我们认为“我”所指称的东西。正如他所写到的，这是“那样的一些情形中的一个，就是其中我们有关我们的一个概念的所指的那些最重要的信念中的一些可能是虚假的，但是并非必然得出没有这样的东西存在”。

我们应当接受这个观点吗？内格尔认为我们不是独立于我们的大脑、躯体和经验而存在的实体。而且看起来他认为，如果“我”这个词并非指称我的大脑，那么它也不会有什么别的好指称的。我的大脑必定是我的经验的主体，因为在缺乏笛卡儿式的自我的情况下，没有什么别的会是我的经验的主体。于是，在否定我们是独立存在的实体之后，他问（1）“为什么不紧跟帕菲特，而且放弃自我与心理性的东西的那个主体的同一呢……？”他的回答是说（2）它纵然不是我们倾向于认为的那类实体，但“那个实际上的主体是重要的”。（1）假定，按

照我所辩护的那个还原论者的观点，我们不再认为是有经验的主体。（2）假定，经验的主体是大脑。

我对这两个假定都否定。按照我所辩护的那个还原论者的观点，个人不是独立存在的实体。一个人的存在正是涉及他的大脑和躯体的存在，以及他的行动的实施和他的精神状态、精神事件的发生。尽管个人不是独立存在的实体，但是他们存在。而一个人是不同于其大脑或躯体以及其各种各样的经验的一个实体。一个人是具有一个大脑和躯体，以及具有不同经验的一个实体。我用"我"这个词指称我自己，一个特定的个人，或经验的主体。而且我不是我的大脑。

回到休谟的类比可能有所帮助。在关于国家方面我们可以是还原论者，但是仍然认为国家存在，而且能够被指称。一个国家不是一个独立存在的实体，不是某种不同于其公民和其公民所居住的土地的东西。一个国家的存在恰好寓于在其领土上以各种各样的方式共同行动的其公民的存在之中。尽管这是一个国家存在之为存在的全部，但是我们能够指称诸国家，并主张它们存在。于是我们能够确实地主张说法国存在，而且法国于1939年对德国宣战。相反，不存在什么被称作鲁里坦尼亚浪漫国的国家。就人们而言我们能够做出相同的主张。某些人存在，而且能够被指称，而其他一些人则根本不存在，而且不能够被指称。我和托马斯·内格尔是两个存在的人，而且能够被指称。但是我们却不能够指称我的非存在的罗马祖先狄奥多里库斯·珀费克图斯。

我的第二个主张在这个讨论中具有特别的重要性。当我们用"法国"这个词指称一个国家的时候，我们不是在指称某种不是国家的东西。我们不是在指称这个国家的政府，或者指称其公民，抑或其领土。这一点可以表明如下。如果"法国"指称法国政府的话，那么倘若法国政府辞职而且具有一段无政府时期的话，法国就会不再存在了。但这是虚假的。在国家没有政府的那些时期，国家继续存在。同样，如果在1939年"法国"指称那时的那些法国公民的话，那么当这些公民

不再存在的时候法国就会不再存在了。这也是虚假的。而且如果“法国”指称这些公民的话，那么一定是这些公民向德国宣的战。这也是虚假的。“法国”这个词有一个用法，它不是指称国民（nation），而是指称国土（country），或者这个国家的领土。当我们主张说法国非常美丽的时候，我们是在指称其土地和建筑。但是按照另一个用法，“法国”指称国民而不是领土。如果“法国”指称法国领土的话，那么除非那片领土不再存在，否则法国就不会不再存在。这也是虚假的。一度曾为普鲁士国家的那片领土仍然存在。但是普鲁士已经不复存在。

正如国家这个例子所表明的，我们能够指称某物，尽管这个某物不是一个独立存在的实体。而且在指称这样的一个事物的时候，我们不是在指称涉及在它的存在之中的其他各种各样的实体。如果我们在关于个人的问题上是还原论者，那么关于我们对于“我”一词的用法我们也能够做出相似的主张。这词能够指称一个人，或者经验主体，纵使一个人不是一个独立存在的实体。而且当我们裁定一个人不是一个独立存在的实体的时候，我们也不是要被迫得出一个人必定要么是他的大脑、要么是他的整个身体这样的结论。虽然国家不是独立存在的实体，我们也不是要被迫得出一个国家必定要么是其政府、要么是其公民、要么是其领土、要么是这三者的结论。一个国家不是这三者中的任何一个。而且我们能够指称诸国家。相似地，我们不是要被迫得出一个人是他的大脑或者他的整个身体的结论。而且我们能够指称诸个人。[3]

我同意内格尔有关我们绝大多数人在“我”这个词所意图的所指方面抱有虚假信念的观点。我们绝大多数人认为我们是些实体，其连续存在必定是要么全有要么全无。可能有这样的诘难，就是说，既然没有什么这样的实体，我们应该得出结论：正如我们绝大多数人所使用的那样，“我”这个词未能指称，就像“宙斯”未能指称一样。我能够像内格尔一样拒斥这个主张。“宙斯”之所以并未指称是因为一种云太不像一个神。按照那个还原论者的观点来说的个人不同于按照非还原论

者的观点来说的个人。但是两者之间比神和云之间要更加相似得多。因而我能够主张说个人存在。而且既然个人存在，那么，尽管与我们倾向于认为的方式不同，但是我们指称个人的尝试能够宣称是取得成功的。像内格尔一样，我们能够主张说，这是那些情形中的一个，其中我们对我们所意图的所指抱有虚假信念，但并非必然得出我们并不是在指称这个事物。

既然我是一个个人，存在着，那么我看起来就是我使用“我”这个词所指称的东西的最佳候选。内格尔可能像下面这样做出回应。的确，我们绝大多数人认为我们是独立存在的实体。按照那个还原论者的观点，这个信念却是虚假的。但是当这个信念应用于一个人的大脑的时候，则会是真的。可能有人宣称这使一个人的大脑成为他使用“我”这个词所指称的东西的一个更好的候选。

这个主张中含有某种东西，但是我认为还不够。如果我使用“X”这个词力图指称被称作X的那个对象，而且X存在，那么自然而然的假定便是我确实指涉X。如果不仅X缺乏我们认为它拥有的绝大多数属性为真，而且别的某个对象Y具有这些属性中的绝大多数也为真，那么这个假定可能得不到正当证明。这可能证明这样一个主张是正当的，那就是，尽管我是在指称X，但我事实上却是在指称Y。当某个人使用“我”这个词的时候，他所意图的所指——他在力图指称的东西——是他自己。这个人可能相信他是一个独立存在的实体，不同于他的大脑和躯体。那么正如内格尔所主张的，这样的一点会是真的，那就是，这个人所意图的所指缺乏他认为它所拥有的那些属性中的一个，而另一个实体——他的大脑——却具有这个属性，即作为一个独立存在的实体（当然并非不同于它本身）。但是一个人的大脑并不具有我们绝大多数人所相信的我们所具有的那绝大多数属性。例如，它并不具有必定要么全有要么全无的一种连续存在；它也不是不可分的。当我们绝大多数人使用“我”这个词的时候，我们的大脑与我们信以为是我们所意图的所指的东西并不非常相似。这算作不利于这样一个主张，那就

是，当我们使用“我”的时候，我们事实上是在指称我们的大脑。算作有利于这个主张的是这样一点，即我们的大脑具有我们误以为我们具有的那些属性中的一个：作为独立存在的实体。但是我认为这并不足以证明内格尔的主张是正当的。它不足以证明我们是在指称我们所意图的所指这个自然而然的回答是虚假的。当我们使用“我”的时候，我们所力图指称的不是我们的大脑而是我们自己。我们的大脑具有我们误以为我们自己具有的一个属性；但是这不足以表明，当我们力图指称我们自己的时候，我们未能做到。我们能够保留我们的这样一个自然信念，就是我们能够指称我们自己。

内格尔的第二个论证诉诸他有关“我”这个词的意义的另一个主张。这是这样一个主张，就是说，在使用“我”的时候，我们意在指称无论什么解释我们的心理连续性的东西。我认为我们能够拒斥这个主张。在内格尔的论证之间存在一种对照。各自涉及一个有关意义的主张和一个有关事实的主张。我接受那个有关我使用“我”意在指称我的经验的那个主体的主张。但是我已否定了我的经验的主体是我的大脑这一主张。我把这个主张认作内格尔的观点，我接受他的关于意义的主张，但是否定他的关于事实的主张。在考虑他的另一个论证的时候，我接受他的关于事实的主张。解释我的心理连续性的东西就是我的大脑的连续存在。但是我否定他的关于意义的主张。我否定那个有关我使用“我”这个词意在指称无论什么解释我的心理连续性的主张。

内格尔的第三个论证诉诸一个想象性的事例，其中在他看来重要的是他的大脑的存在。他描述了一个像电子传输那样的事例。在这个事例中，许多人会接受内格尔的主张。他们会认为，重要的是他们的大脑的存在。

我现在将描述有关这一点更加难以置信的两个事例。首先谨记的是，即便一个对象的所有成分都被替代，它也可能继续存在。标准例

子是那条木船的例子，木船在每一次航行之后，它的一块船板就会被替换掉。我们可能认为大脑也是如此。我们已经知晓，我们躯体的其他部分的细胞全都渐次被替代。即便这一点对于我们的大脑而言不为真，但是它本来可能为真。即便我们的大脑像我们躯体的其他部分一样，其全部成分被自然地、渐渐地替代，我们的大脑可能本来会继续存在。而且我们可能认为，如果我们自己造成了这样的一种渐进的替代的话，那么我们的大脑会继续存在。我将在此假定这是真的。既然我们认为一个人躯体的其余部分的成分尽管被逐渐替代它却确实继续存在，那么我们为什么对我们的大脑应当持有不同的观点?

下面假设我需要手术。我的脑细胞全都具有一种缺陷，而这个缺陷在时间过程中会是致命的。但是一个外科医生能够替换掉所有这些细胞。他能够填入新细胞，这些细胞除了没有缺陷之外，完全是那些现存细胞的精确复制品。我们可以区分出两种事例。

在事例一中，那位外科医生实施100次手术。各次手术中，他去除我的大脑的1／100，而且置入这个部分的一个复制品。在事例二中，那位外科医生遵循一种与之不同的程序。他先是去除我的大脑的所有部分，然后置入它们的所有复制品。

在这两种事例之间存在一种真正的差异。在事例一中，我的大脑的那些新部分各自有一段时间与我的大脑的其余部分相结合。这能够使各个新部分成为我的大脑的一部分。当第一个新部分被置入并且与那些旧有的部分相结合的时候，它就像那些旧有的部分一样多地赢得了作为我的大脑的一部分的资格。当第二个新部分被置入的时候，它也成为我的大脑的一部分。就每一个新部分而言这一点都成立，因为都有一段这个部分与其时被算作我的大脑的其余部分的东西相结合的时间。

在事例二中，情况则不同。各个新部分没有什么与我的大脑的其余部分相结合的时间。因此，那些新部分并不算作我的大脑的一些部分。我的大脑不再存在。

有关一个俱乐部的存在或许有某种相似的为真之处。考虑一下一个成员人数限于50人的俱乐部。所有现存的成员全都想要退出。另外有50人想要加入这个俱乐部。但有一条规则规定，除非有49个现存的成员在，否则一个人都不能够被接收为新成员。因为这条规定，情况的发生只有像事例一那样，这个俱乐部才继续存在。所发生的情况必须是这样的。一个成员退出，一个新成员被接受。另一个成员退出，另一个新成员被接受。第三个成员退出，又有一个新成员被接受。在这个系列的最后，这个俱乐部全都是新的成员，但它会仍然存在。假设相反所发生的情况像事例二一样。老成员全都退出。因为这条规定，现在不能够接受新成员。该俱乐部不再存在。

回到事例一和事例二。我是在假定，大脑可能通过一个逐渐替代的过程而变成由新的一些成分所构成。按照这个假定，显然在事例一中我的大脑继续存在，而在事例二中它并不如此。按照这个观点，事例一带给我生命，而事例二带给我死亡。

这令人信服吗？尽管这两个事例之间存在一种真实的差异，但是比内格尔头脑中所想的差异要小一些。他考虑过一个事例，其中他的大脑会被毁灭，而一个复制品被创造出来。而且他把这个事例与通常的存活做了比较，在该事例中他的大脑继续具有所有的那同一些现存的细胞。

在我所举出的一对事例中，差异更加小一些。在我所举出的两个事例中，后来都将有一个人，其大脑除了没有那些缺陷之外将确切地像我现在的大脑。结果，这个人将在心理上与我相连续。而且在这两个事例中，这个人的大脑都将由一些完全相同的新成分所构成，这些成分各自是我的大脑的某个部分的复制品。两种事例之间的差异只在这些新的部分的置入方式。这是一个移除和置入顺序方面的差异。在

事例一中，那位外科医生在移除和置入之间交替进行。在事例二中，他在一揽子置入之前做了一揽子移除。

这能够是生死之间的差异吗？我的命运取决于移除和置入顺序上的这个差异吗？就我的幸存而言，那些新部分是否有一个与旧有的部分相结合的时间就那么重要吗？如果产生出某个进一步的事实的话，这会造成不同。如果我的幸存就像某种只有祭司通过仪式性的触摸才能够传给另一个人的神圣力量的话，就会如此。但是，并没有这样的进一步事实。有的只有这样一个事实，即如果那些新部分有一个与那些旧部分结合的时间，我们把那个结果产生的大脑描述为那同一个大脑。如果那些新部分并不如此结合的话，我们把那个结果产生的大脑描述为一个与之不同的大脑。

内格尔并不认为他的观点的那些根据是决定性的。而且他承认，“难以内化一个我自身等同于我的大脑的概念”。他之所以接受他的观点，部分是因为在他所考虑的那一对事例中，他的幸存在他看来取决于他的大脑是否继续存在。在我所举的那一对事例中，所发生的事情的差异要小得多。如果他考虑过这些事例的话，内格尔也许会改变他的观点。他既提到（1）同一性是重要的，又提到（2）他是他的大脑。但是他承认（2）难以接受。难以把自身想作自身的大脑。当我考虑事例一和事例二的时候，我发现既相信（1）又相信（2）是不可能的。我不能够相信，对我的幸存而言重要的是我的大脑的那些部分的复制品在某段时间是否会以这两种方式之一来置入。我不能够相信，如果那位外科医生交替地移除和置入，那么这就像通常的幸存一样好；而如果他在一次性置入之前做了一次性移除，那么这就几乎像通常的死亡一样糟。如果两个事例之间的这个差异并不是真正重要的，那么就有两个可供的选择。要么同一性并不重要，要么我并不是我的大脑。

第一种选择得到我出现分裂的那个想象事例的支持。在这个事例中，结果产生的两个人各自具有我的半个大脑。而且没有复制。这两

个半脑将由我的现存的大脑细胞构成。正如我已经论证的，非常难以相信的一点是，我应当视分裂等同于死亡。我与结果产生的各人之间的关系包含正常幸存所会要求的一切。纵使我是我的大脑，这一点仍然保持为真。我的大脑的各个半脑将继续存在，而且继续支持有意识的生活。具有我的半个大脑的各人，看起来将记得我的整个生活，而且以所有的其他方式在心理上与我相连续。如果我是我的大脑，那么这就不是一个其中我因为我的大脑不再存在而死亡的事例。我的大脑继续存在，而且因为它分裂了，它对生活的支持更加丰富。它所支持的不是一个生活，而是两个生活。

假设内格尔同意这样一点，就是我与结果产生的各个人之间的关系包含重要的东西。那么他会辩护他的有关同一性重要的假定吗？我们已经看到他会的，他的手法是对我们的个人概念可笑的歪曲。我假定内格尔会拒斥这些歪曲。他会同意，在我分裂之后，我将不再会一个人活着。如果他还同意我与结果产生的各人之间的关系就像幸存一样好，他就必须放弃同一性重要这一假定。没有这个假定，我就不是要被迫得出复制会像死亡一样糟的结论。我能够同意我的复制品不是我，但是主张我与他的关系包含根本重要的东西。我能够主张重要的是带有任何原因的心理联系性和／或者连续性。

内格尔可能回应如下。他或许同意，有一种特殊的情况，其中同一性不是重要的。这就是其中未来的两个人各自具有我的半个大脑的那个情形。但是此处重要的仅是我的分裂了的大脑的连续存在。在别的每一情形中，个人同一性是重要的。

这个回应可能具有某种力量。但是如果我们认为同一性是重要的，那么认为同一性总是重要的就是自然而然的。如果我们承认有一个例外，那么就可能难以证明拒斥其他的是正当的。鉴于我所举的事例一和事例二之间的微小差异，所以我们在此也能够主张说，同一性并不是重要的。如果这个主张在我分裂的那个事例中得到正当证明，那么此处为什么不能够得到正当证明？难以相信我的命运取决于这些

情形之间的差异。与内格尔所考虑的那对事例不一样的是，这对事例提示着我不是我的大脑。

　　我尝试着回答了内格尔的论证。我的那些回答并不表明他的观点是虚假的。但是我认为它们表明，我们能够可辩护地持一种与之不同的观点。问题仍然悬而未决。这就是为什么我在第98节中对内格尔的观点提出一个相当不同的回应的原因。

---

[1]参见Kripke通篇。这个观点也得到H·普特南数篇文章的发展。

[2]我们在指称的时常取决于我们某个语言部分的因果历史。但是情况并非总是如此；对于此一种类进行因果考虑看来与内格尔所提出的那个特定观点并不相关。

[3]参见第三编注释中所引述的Kripke的话。

E 最近连续者图式

诺奇克就历时同一性的所有判断提出一个一般性的观点。要成为像某个过去事物相同的事物就是要成为那个事物的最近连续者。诺奇克的观点是还原论者的观点。他主张在某个过去的人与各种各样的当前的人们之间能够存在各种连续性。我们判定是这个过去的人的那个当前的人就是与这个过去的人具有最大连续性的那个当前的人。按照这个观点，历时的个人同一性这一事实就寓于保有这样的诸连续性之中。它并不涉及任何进一步的事实。而且诺奇克明确比较了个人同一性与一个特定的哲学家群体维也纳学派的历时的同一性。这就像我对一个俱乐部的历史所作的比较。

尽管诺奇克的观点在这些方面是还原论者的观点，但是他拒斥我所辩护的那个还原论的版本。他写道：

> 接近那些程度各异的复杂关系的一个纷乱领域的哲学进路之一，不是力图把它们强塞入有些随意的分类架，而是满足于辨认并勾勒那些处于底层的复杂关系。就个人同一性而言，可能可以说，借助多样的处于根基的关系和事件，诸如躯体的连续性、心理的相似性、分裂或聚合，未来的诸自我将与我们-现在具有程度不一的接近性；有待告知的真正的和整个的真相就是这些处于根基的现象的存在和轮廓。为什么要把范畴化——最近连续者图式就是一个——加诸这个复杂性？

他然后明白无误地拒斥我的观点，他继续写道：

> 不过，根基层本身也将产生类似的难题。例如，当人体的构成神经细胞的那些特定的分子以及所有不是神经细胞的细胞都历时地被替换的时候，它在什么方面还是那同一个躯体？我们只是应当再次谈及作为这个水平根基的那些复杂关系吗？我们不能够凭借把我们自己限制在这些处于根基的关系的完全的复杂性……避免那个最近连续者图式，或者别的什么范畴化。最终我们会被推到一个最近连续者图式或某种在这样那样的层面上与之相似的东西……如果因为必须，在某个层面上使用该图式成为合法的，那么我们为什么不就从它开始呢？[1]

这个争论是不必要的。我并不否认我们对许多不同种类事物历时的同一性作出判断。而且我把诺奇克的最近连续者图式作为我们如何作出许多这样的判断的一个说法来接受。我的主张是这样的。既然历时的个人同一性就寓于保有某些别的关系之中，那么重要的不是同一性而是这些别的关系中的一些。而且同一性的逻辑并不总是与什么重要相契合。当重要的东西呈现为一种分支形式的时候，或者在一些中间程度上成立的时候，就不能够令人信服地使同一性的判断与重要的东西相一致。在这些情况下，我们不应当使用最近连续者图式来试图取得这样的一致性。正如诺奇克所写到的，那么我们就会把重要的东西强行"塞入有些随意的分类架"。在这些情况下，我们只应当描述这些别的关系在什么方面和在什么程度上成立。然后我们应当努力判定这些关系在不同的方面重要的程度。上文所引述的诺奇克对这个观点的诘难看起来是说，我们不能够避免对同一性作出一些判断。但是，只有当我们添加这样的一个主张的时候，就是主张倘若我们通过对同一性作出判断或者否定有关同一性的判断，那么我们必须以这个方式描述所有情形的时候，这才是一个诘难。我看不出有什么理由来接受这个主张。

在诺奇克有关什么重要的主张中，看来他再次拒斥还原论。他写到，按照他的观点，"我将同等关心我的最近连续者，无论其接近程度碰巧如何（倘若是足够接近的）"。如果他考虑过我的联合谱系的话，诺奇克就可能会收回他的这个主张。在这个谱系的那些中间情形中，会有结果产生的一个人，他会有我的大脑和躯体的一定比例的细胞，而且会与现在的我在心理上有某些方面的连续性。但是这个结果产生的人还会有许多的、不同的细胞，他或者她还会在许多方面与格丽泰·嘉宝在心理上具有连续性。在这个谱系的远端，这个未来的人不会在什么方面与我相联。如果我接受诺奇克的观点，我就同等地关心这样的一个未来的人，倘若他或者她与我足够近地相关联。我把这个谱系中第一部分中的所有情形都视作像通常的幸存一样好。随着我们沿着

这个谱系移动，那个未来的人与我相关联的程度会越来越不接近。但是我仍然同等地关心这个人，倘若接近的程度足够近。

按照这个观点，我必须决定什么程度的接近算作足够接近。我再次必须在这个谱系上划一条清楚的线。如果我与某个未来人的关系恰好在这条线的近的一边，这个关系就像通常的幸存一样好。如果我与某个未来人的关系恰好在这条线之外，那么我就应当少关心一些。但是第二个情形中的那个未来的人与第一个情形中的那个人几乎没有什么不同。差异只会是更多一些细胞会被替换掉，以及会有某种微小的心理变化，诸如一个想独处的新愿望等。尽管这些是惟有的差异，但是我应当较少关心发生在这第二个人身上的事情。

这个关心模式在我看来是非理性的。是否会有更多的细胞被替代，或者是否会多一个微小的心理变化，何以能够具有这样的重要性？诺奇克的观点把这个谱系当作宛如它在某一点上牵涉到一个非连续性一样。但这是虚假的。既然这个谱系是平滑的，囊括所有程度的连续性，为什么在该谱系第一部分的所有情形中同等地关心，然后却突然地关心少了？只有当同一性是某个进一步的事实，在该谱系的第一部分中完全成立，而且然后突然地未能成立的时候，这才会是合理的。但是诺齐克并不相信有任何这样的进一步事实。

诺奇克提到他的关心模式或许得到辩护的另一个方面。他认为，接受他所称对某物关心的柏拉图式的模式会是合理的。在这个模式中，"我们以世界实现其彼岸的东西的那个方面来看世界，我们看到并且能够对世界所闪现的某种更加精致的东西的闪光作出反应"。即便我们是一些其存在要么全有要么全无的存在者这一点并不为真，我们以这一点仿佛是真的那样来关心我们的同一性也能够是合理的。正如诺奇克所承认的，这涉及到"一种对现实性的不切实际的高估，一种透过柏拉图式的眼镜对现实性的看法"。替代选择是作出"一种对事物的更加切合实际的评估，按照事物本身的样子来看事物"。诺奇克诘难这个

替代选择，认为它"使一个人成为现实世界的阶下囚或受害者，受到它的有缺陷的方式的限制，受到它碰巧所是的样子的限制……"。[2]

这是对诺奇克的关心模式的一种充分的辩护吗？他的关心不是相应于他的生活的现实真相而是他所期望真相本来所是的那样，这一点会是合理的吗？有鉴于理论的和实践的合理性的区分，诺奇克的关心模式是可辩护的。因此我收回我上述的诘难。再也不存在什么争议了。如果诺奇克对现实的反应不是按照现实的本来样子，而是按照他但愿它是的样子，这在理论上是非理性的。但是如果这种希望的思维是令人更加深深地满意的话，那么就他而言力图使他自己以这一方式在理论上非理性这一点在实践上会是合理的。

---

[1]Nozick（3），第60—61页。

[2]Nozick（3），第67—68页。

F 社会贴现率

根据一个社会贴现率[1]，未来事件的当前道德重要性，尤其是盈亏，每年以n速率递减。两个常用的速率是5%和10%。对一种SDR而言，我并无争议。这种SDR应用于基于通货膨胀假定的、以货币度量的盈亏。但是许多经济学家把一种SDR应用于以其发生规模度量的盈亏。文书中我所举的例子不是想象出来的。下面一点已经被严肃地提了出来，那就是，当估算核废料的处置风险的时候，我们应当把一种SDR应用于未来的死亡。更宽泛一些说，诸社会贴现率不仅已经被应用于货币的盈亏，而且被应用于经济学家们所称的、未来的人们将享受到的实际功利。这就是我所讨论的那种SDR。这种SDR为什么已经被认为是得到正当证明的呢？我意识到有六大论证：

民主论证：许多人对更加遥远的未来关心得少一些。一些学者主张，如果就某个民主国家中的绝大多数成年公民而言这一点为真的话，那么这个国家的政府就应该应用一种社会贴现率。如果其选民的确对更加遥远的未来关心得少一些的话，一个民主的政府就应该应用社会贴现率。未能做到这一点就会是家长式统治的、集权主义的。用一位作者的话来说，该政府的决策应当"只反应当前的各个个人的优先选择"[2]。我们可以忽略这个论证。存在的问题有两个：

（1）作为一个共同体，我们可能使用一种社会贴现率吗？我们以每年n%的某个速率对我们的社会政策的更遥远的后果递减地关心，这在道德上得到了正当证明吗？

（2）如果我们的共同体中的绝大多数人对问题（1）给出肯定的回答，我们的政府应该不理会这个大多数人的观点吗？

民主论证只适用于问题（2）。对我们所关心的问题（1）来说，这个论证是不相关的。

这一点或许可以这样来表达。一个民主主义者相信某些特定的宪政安排。这些安排提供了他对问题（2）的答案。他对民主的承诺如何才会给予他对问题（1）的答案？只有当他假定大多数人所想要的

或者信以为真的就必定是正确的这一点的时候。但是没有什么合乎情理的民主主义者假定这一点。假设大多数人想要发动一场侵略战争，毫不在乎那些无辜的外国人遭受生灵涂炭。这不会表明他们不顾生灵涂炭就是对的。同样，即便我们绝大多数人确实对我们的社会政策的更遥远的后果关心得少一些，而且相信这样的较少关心在道德上是得到了正当证明的，这也不能够表明它就是得到了正当证明的。无论我们绝大多数人想要或者相信的是什么，这个道德问题仍然悬而未决。

可能有这样的诘难：“在某些情况下，这并不是一个道德问题。假设，在某场公投中，我们投票赞成一项只将影响到我们自己的社会政策。而且假设，因为我们较少关心将在以后发生在我们身上的事情，我们投票赞成一项现在带给我们获益、但代价是以后承受更大负担的政策。这项政策是有悖于我们的利益的。但是，既然这项政策只将影响到我们自己，我们投票赞成这项政策就不能够是在错误地行动。我们充其量会是在非理性地行动。”

按照我们绝大多数人所采纳的那些假定，这些主张确实提供了对该社会贴现率的某种辩护。但是这种辩护很少适用。绝大多数社会政策不仅影响到我们自己，而且将影响到我们的子孙后代。如果我们投票赞成一项将有悖于我们子孙后代利益的政策，我们绝大多数人会承认，在某种程度上，这在道德上会是错误的。相似的评述适用于那些尚未出生的人们的利益。我们应该赋予这些人的利益多大的权重是一个道德问题。另一个诘难是说投票很少是全体一致的。如果某项政策将有悖于我们的利益，它将很有可能是有悖于那些投反对票的少数人的利益。那么，投赞成票的大多数人的行动就会是有悖于这些少数人的利益。这是道德批判的另一个根据。

正如这些评述所表明的，极少会有该社会贴现率并不引发道德问题的情形。如果我先前的主张之一是得到正当证明的，就不会出现这样的情形。我论证过，如果我们对将在以后发生在我们身上的事情较少关心，而且行动从而有悖于我们的利益，那么这可能并不是非理性

的。但是这样的一些行动有遭受批判的空间。我主张的是，我们应当认为它们在道德上是错的。如果是这样，那么该社会贴现率总会引发道德问题。

概率论证：常有主张说，我们应当对那些更遥远的后果打折扣，因为它们较不可能发生。

同样存在两个问题：

（1） 当一个预测应用于更遥远的未来的时候，它较不可能是正确的吗?

（2） 如果某个预测是正确的，我们可以因为它应用于更遥远的未来而给予它较少的权重吗?

对（1）的回答经常是肯定的。但是这并不为对（2）作出肯定的回答提供什么论证。

假设我们是在决定是停止还是增加我们对核能源的利用。我们在借助估算预见到的核辐射泄漏所导致的死亡人数来考量可能的事故。在一场小事故中，在我们永远不会知晓这场事故造成那些特定的人死亡的意义上，这样死亡的人可能全都只保持一种统计学上的地位。当考量可能的事故的时候，我们必须思考到遥远的未来，因为一些放射性元素的危险性保持长达数千年。根据一个5%的社会贴现率，统计学上的明年1个人的死亡算来大于400年后10亿人的死亡。较之于造成这个单一的人的死亡，我们所选择的政策是否造成这10亿人的死亡则在道德上较不重要。这个结论是令人发指的。这10亿人会在更遥远的未来被夺去生命。但是这不能够证明这样的一个主张是正当的，就是说，较之于杀死该单一的人，如果我们相反杀死这10亿人，那么我们会是在较不那么糟地行动。概率论证并不导致这个结论。它充其量会导致一个与之不同的结论。我们明白，如果核辐射明年泄漏，我们将没有什么充足的防护。我们可能认为，历经下面的4个世纪，将有某种应对措施或者治愈手段被发明出来。于是我们可能认为，如果核辐

射在400年之后泄漏，将更少可能造成死亡。如果我们是非常乐观的，我们可能认为可能性将少10亿倍。这会是以10亿为因数对400年后的死亡人数打折扣的一个不同的理由。我们不会是在做这样的一个令人发指的主张，就是说，如果我们确实造成这样的死亡，那么各个死亡的人将比明年死亡的一个单一的人在道德上的重要性少10亿倍。相反我们会是在主张，造成这些更遥远的死亡的可能性小10亿倍。这是为什么，以我们的观点看来，我们几乎不需要关心400年之后核辐射泄漏的原委。如果我们主张这样的死亡可能性小10亿倍正确的话，我们的结论就会是得到正当证明的了。并不发生的死亡，无论是现在还是400年之后，都并不重要。

这个例子例证了一个一般性的要点。我们应该对那些更有可能虚假的预测打折扣。且让我们称之为概率贴现率。有关更遥远未来的预测更有可能虚假。所以时间性的和概率性的这两种贴现率，大致是关联的。但它们是相当不同的。因此，因时间而不是因概率打折扣是一个错误。有一个诘难就说，这误述了我们的道德观点。它使我们不是主张更遥远的那些糟糕后果更少可能，而是主张它们并不重要。这并不是我们的真实观点。一个更大的诘难是说，这两个贴现率并不总是相契合的。有关更遥远的未来的那些预测以每年n%的某个速率并非越来越不可能为真。当应用于遥远的未来的时候，许多预测的确更大可能为真。如果我们因时间而不因概率来打折扣，如是我们可能被引导到得出那些即便按照我们自己的假定来看也是错误的结论。

机会成本论证：有时候更早一点得到某个收益要更好一些，因为这个收益之后能够用来产生一些进一步的收益。如果一项投资明年产生回报，这将比10年之后相同的回报更值得，倘若这个更早一些的回报能够在这10年间获利地再投资的话。当我们加上来自这个再投资的额外收益的时候，那么获益总额将更大一些。相似的论证涵盖某些种类的成本。如此一来某些获益的延宕就牵涉到机会成本，而且反之亦然。

有时这被认为是为一种社会贴现率提供正当证明。但是该正当证明是失败的，而且是出于相同的两个理由。某些机会成本确实历时地

增长。但是如果我们因时间而打折扣，而不只是加进这些额外的成本，我们将错误表达我们的道德推理。更为重要的是，我们可能被导入歧途。

考虑一下那些不是被再投资而是被消费的获益。当后来得到这样的获益的时候，这可能并不涉及到什么机会成本。假设我们是在决定是否要建设某个机场。这会毁掉某片美丽的乡村景色。我们会失掉享受这个自然美景的利益。如果我们不建设这个拟议中的机场，我们以及我们的子孙会在未来的每一年享受到这个利益。根据一种社会贴现率，后来一些年份中的那些获益要比明年的那个获益要少算许多。诉诸机会成本如何会证明这一点正当？明年得到的那个获益——我们对这个自然美景的享受——不能够被有利可图地再投资。

该论证也不能够应用于那些仅"被消费的"成本。假设我们知道，如果我们采纳某项政策，那么将有某种造成一些基因畸形的风险。该论证不能够表明，明年的一个基因畸形应当算作20年后一个基因畸形的10倍。所会主张的充其量不过是这样一点。我们或许断定，对各个受到这样影响的孩子而言，总额为k元的一大笔钱会提供足够的补偿。假如我们将会提供这样的补偿的话，用于确保这一点的当前成本会因明年所造成的一个畸形而大得多。我们现在就得留出几乎足额的k元的钱。这个总额的区区1／10，如果现在留存下来并且获益地投资的话，或许20年之后就会等值于k元。这提供了现在较少关心我们或许在更遥远的未来所造成的那些畸形的一个理由。但是该理由并不是这样的畸形较不重要。该理由是它现在会给我们造成的成本只有用以确保这样的一些畸形发生的时候我们能够提供补偿这一点所用成本的1／10。这是一个至关重要的差异。假设我们知道，我们事实上将不会提供补偿。例如，假如我们不能够确认我们的政策已造成了那些特定的基因畸形，就会如此。这消除了我们现在较少关心以后一些年份中的畸形的那个理由。如果我们将不为这样的一些畸形提供补偿，那么就那些后来的畸形而言，现在用以确保我们本可提供补偿这一点的成

本本会更便宜一些这一事实就成为不相关的。但是如果这个事实已经导致我们接受一种社会贴现率，那么我们可能未能注意到它什么时候变得不相关。我们可能被导致假定，即便在没有什么补偿的时候，20年之后的畸形的重要程度只有下一年的那些畸形的重要程度的1／10。

这里还有另一个对该论证的诘难。在某些特定时期，投资并不带来什么回报。当此点为真的时候，机会成本论证就不适用了。该社会贴现率又未能与任何重要的东西相关联。

这些简短的话忽略了许多争论点。就对一种社会贴现率的各种各样的论证而言，所诉诸的机会成本是最难以估定的。但我认为主要问题是简单的。使用一种社会贴现率来更方便、更漂亮地计算那些机会成本的方法可能有数种。但是那些靠这样的计算所得出的最后结果本可以用一种时间上中立的方式加以重新表达。当表述未来的那些政策的后果的时候，经济学家们会用一种并不使用贴现率的方式来陈述不同的时候会有什么未来的获益和成本。那些诉诸机会成本论证本来能够用这样的一些术语来加以充分表达。我相信，就我们需要作出决定的任何重要问题而言，这会是对那些可供选择的一个更好（因为更少误导）的表述。它会使之更容易达到正确的决定。

我们的子孙将更好地论证：如果我们假定我们的子孙将比我们现在更好，那么有两个令人信服的、对于那些我们给予他们和强加给他们的获益和成本打折扣的论证。如果我们以那种出于通货膨胀而调整过的货币核算来度量获益和成本，那么我们能够诉诸货币渐减的边缘效用。同样的财富增长一般而言带给那些境况更好的人们的利益要更小一些。我们还可以诉诸一种分配原则。给予境况更好的那些人们的一个同等大的利益可能被认为在道德上的重要性更小。

尽管这是两个好的论证，但是并未证明社会贴现率是正当的。对这些未来的利益打折扣的根据不是它们出现在更遥远的未来，而是它们将落在那些境况更好的人们身上。此处，就像在别处一样，我们应当说出我们所意味的。而且相互关联又不完美。我们的一些子孙可能

并不比我们现在境况更好。如果他们并不更好的话，刚才给出的论证也就不能适用了。

> 过度牺牲论证：一个典型的陈述是这样的："出于一些理论性的理由我们显然需要某种贴现率。否则，任何遥远未来的微小利益的增加都可能要求一定量的当前的牺牲，因为最终那些利益会超过该成本。"

同样的一些诘难适用于此。如果这就是我们为什么采纳社会贴现率的原委，那么我们将误述我们之所信。我们的信念不是未来利益的重要性稳步减少，而是没有哪一代人能够在道德上被要求为未来的那些世代的人之故而作出多出某些种类之外的牺牲。如果这就是我们所相信的，那么这就是应当影响我们决定的东西。

我们可能具有另一个信念。如果我们瞄准历时的最大的净利益总额，那么这可能要求在不同的世代之间进行非常不均的分配。功利主义者们会主张说，鉴于那些切合实际的假定，这不会是真的。但是假定这是真的，那么我们可能希望避免得出应该有这样一种不均的分配这一结论。而且，如果我们对以后的那些利益打折扣，我们在一些情形中能够避免这个结论。然而，正如罗尔斯所指出的，这是避免这个结论的一种错误方法。如果我们并不认为应该有这样的不平等，那么我们就不应当只是瞄准最大的净利益总额。我们还应当有第二个道德目标：那些利益在不同的世代之间得到公平分享。在我们的功利原则之上，我们还应当追加一个公平分配的原则。这更准确地表述我们的真实观点。而且它撤除了我们对以后的那些利益打折扣的理由。

假如相反我们通过采纳一种社会贴现率来表达我们的观点，那么我们会被无谓地引导到一些不令人信服的结论。假设，以对我们自己现在而言的同一成本，我们要么能避免较近的未来中的一场小灾难，要么能避免更遥远未来中的一场大灾难。既然避免那场大灾难不会涉及到什么额外成本，那么过度牺牲论证就不能适用。但是如果我们采

用那个论证来证明一种贴现率的正当性，那么我们将被引导得出那场更大的灾难较不值得避免的结论。

特殊关系论证：一些功利主义者主张，各人应当给予人人的利益以同等的权重。这并不是我们绝大多数人所相信的。根据常识道德，我们应该给予陌生人的利益以某种权重。但是对于某些特定的人我们要么可能要么应该给予某些种类的优先权。于是我们在道德上被允许给予我们自己的利益以某些种类的优先权。我们绝大多数人相信，当助人这一点要求我们做出过大的牺牲的时候，我们没有什么义务去帮助他人。况且有某些特定的人，我们应该对他们的利益给予某种优先权。这些就是我们与之处于特殊关系之中的那些人。于是各人应该把某种优先权给予他的子女、父母、学生、病人、他所代表的那些人的利益或者他的同胞们的利益。

这样的一种观点自然适用于我们的行动对于未来的世代所产生的后果。我们的直接后继者将是我们的子女。根据常识，我们应该给予他们的福祉以特殊的权重。尽管程度有所降低，但是关于我们的子女的子女，我们同样会这么认为。

相似的一些主张在共同体的层面上看起来也令人信服。我们认为，我们的政府应该特别关心它自己的公民的利益。主张应该特别关心它的公民们的未来子女，以及在较小的程度上关心他们的孙辈，这一点会是自然而然的。

这样的一些主张或许支持一种新的贴现率。在此我们可能不是在出于时间本身来打折扣，而是出于血缘关系的程度。但是至少这两种关系不可能是极端分离的。我们的孙辈不可能全都在我们的子女辈之前出生。既然关联关系在此更加牢靠，那么我们可能受到运用一种标准的贴现率的诱惑。在此我还相信，这会得到正当证明。就一件事情而言，按照任何贴现率，较为遥远的后果总是算得较少举足轻重。但是与亲缘关系相关的一种贴现率在某点上应当不再适用——或者，为了避免非连续性，应当渐趋接近某个高于零点的水平。我们应该给予我们的行动对那些纯粹陌生的人的后果以某些权重。我们不应该给予我们的行动对我们自己的后代的影响后果以更少权重。

这个贴现率也不应当应用于所有种类的后果。请考虑一下这样一个对比。也许美国政府一般应该给予其公民的福祉以优先权。但是这并不适用于施加严重伤害的情况。假设该政府决定恢复大气层核试验。如果预测到结果产生的辐射尘会导致数人死亡，应当对外国人的死亡打折扣吗？从而应当把试验转移到印度洋吗？我认为，在这样的情况下，这种特殊关系并不造成什么道德差异。关于我们施加给我们遥远的后继者们的那些伤害，我们应当持相同观点。

至此我已经讨论了对那种社会贴现率的六大论证。没有哪一个论证是成功的。它们充其量本来能证明把这样的一种比率作为一种粗陋的经验法则来使用的正当性。但是这个法则经常会走入歧途。较少关心我们社会政策的更加遥远的后果在道德上可能经常是允许的。但这决不会是因为这些后果是更加遥远的。而会是因为它们较少有可能发生，或者会是因为它们是对那些比我们境况更好的人们的影响，或者是因为它现在会比给予补偿更加低廉，抑或会是因为我已经给出的其他理由中的一个。所有的这些不同的理由都需要按照它们自己本身的长短来分别加以陈述和判断。如果我们把它们捆绑到一种社会贴现率之中，我们就使我们自己在道德上变得盲目。

时间上的遥远大致与具有道德重要性的那些事实的整个变域相互关联。空间上的遥远也是如此。那些我们身负最大义务的人们，即我们的家人，通常与我们居住在同一屋檐下。我们的居住时常靠近那些我们身负别的特殊义务的人们，诸如我们的客户、学生或者病人。我们的绝大多数同胞比绝大多数外国人住得离我们更近一些。但是没有什么人提出，因为存在这样的相互关联，所以我们应当采纳一种空间贴现率。没有什么人认为，如果我们以每码n%的某个比率递减关心我们行动的长程后果的话，我们在道德上是会得到正当证明的。时间贴现率，我相信，同样没有多少正当证明。

当那些别的论证不适用的时候，我们应当同等关心我们行动的那些可以预见的后果，无论它们将在1年之后发生，100年之后发生，还

是1000年之后发生。这一点意义重大。即便在遥远的未来，一些影响也是可以预测的。核废料的危险可能长达数千年之久。而且我们的某些行动具有永久的影响后果。例如，某个物种、我们大量的环境或者不可替代的文化遗产的毁灭就会是如此。

---

[1]社会贴现率：即下文中的SDR。——译者

[2]Marglin。

G 导致某人存在是否有益于这个人

这个问题一直奇怪地被人们所忽视。于是，在美国参议院“人口增长与美国经济”的一份报告中，出现了这样的主张：“从美国人口的连续增长中不会有什么实质性的获益”。这个报告根本没考虑到，倘若额外的美国人出生的话，这可能有益于这些美国人。

如果我们要捍卫某个有关人口过剩的观点，我们就必须考虑这个问题。如果某个行动是一个具有值得过的生活的人存在原因的一个必然部分，那么这个行动从而有益于这个人吗？我将论证对此作出肯定的回答并非像一些人所主张的那样，是明显错误的。

一些诘难者主张说，生活不能够被判定为比非存在更好或者更糟。但是某种特定种类的生活可以被判定为是好的或是坏的——值得过或不值得过。如果某种特定的生活是好的，那么它就比不存在更好。如果它是糟的，它就比不存在更糟。在判定某个人的生活是值得过的或比不存在更好的时候，我们无需蕴含，假如这个人从来没有存在过的话，对他而言本来会更糟。

此类判断时常是有关某个生活的最后部分而得出的。考虑一下有人正在痛苦地死去，此人已经向人们道过别。这个人可能断定苟延残喘会比死去更糟。要作出这个判断，他无需比较苟延残喘会是什么样子与已经死了会是什么样子。正如威廉斯所写到的，他“可能考虑摆在他面前的是什么，以及决定是要经受还是不要经受”。[1]而且他或许以同样的方式决定他对置于身后的东西是感到高兴还是感到懊悔。他或许断定，在过去的某个点上，假如他知道摆在他面前的是什么的话，他本来会想要或者本来不会想要度过他的余生。他或许如此一来得出他的生活的这些部分比不存在更好或者比不存在更糟的结论。如果这样的一些主张能够适用于一个生活的那些部分，那么，我相信，它们就能够适用于所有的生活。[2]

那些诘难者现在可能诉诸

双态要求：只有当我们导致某人比否则他那时本来会有的境况更好的时候，我们才有益于他。

他们可能说："导致某人存在且具有一种值得过的生活，我们不是在使这个人比他本来会有的处境更好。假如他从来没有存在过的话，这个人本来就不会处境更糟。"

要评估这个判断，我们首先应当问下面这个问题。如果某人现在存在且具有一个值得过的生活，他比假如他死了且不再存在的话现在会有的处境更好？假设答案是肯定的。在应用这个双态要求的时候，我们把已经不再存在算作某人能够其中处境更糟的一个状态。有关从未存在我们为什么就不能够做出同一个主张？我们为什么不能够主张，如果某人现在存在，有一个值得过的生活，那么他就比假如他从来没有存在过的话会有的境况更好一些？从未存在固然不是一个正常状态，但已经不再存在也不是。当我们应用这个双态要求的时候对这两者一视同仁又错在何处？

可能有回应说，当某个人死亡的时候，有一个特定的人，他已经不再存在。我们能够指称这个人。相反，但是没有什么从未存在的特定的人们。我们不能够指称任何一个这样的人。

假如我们是在主张说，我们导致人们从未存在这一点会是在伤害这些人的话，这可能是一个好的回应。但我们是在做一个与之不同的主张。这个主张就是，我们导致某人存在是会有益于这个人。既然这个人确实存在，那么当描述这种备选的时候我们就能够指称这个人。在这一可能的备选中，我们知道谁是那个本来会永不存在的人。在我们正在考虑的那些情形中，已经不再存在和从未存在之间并没有那种所谓的差异。就像我们能够指称现在或许已经不再存在了的那个人一样，我也能够指称或许从来就没有存在过的那个人。我们尚未得到表

明，为什么在应用这个双态要求的时候我们不应当以相同的方式对待这两个状态。

这个双态要求的辩护者或许接着改变他有关死了或者已经不再存在之状态的观点。他可能主张说这并不是一个其中有人能够境况更糟的状态。然后关于从不存在他就能做出同一个主张。

借助这个修正，这个双态要求变得过强。它蕴含着，拯救某人的生命不能够有益于这个人，因为那个被拯救的人的境况并不比假如已经不再存在的话会有的境况更好。在拯救生命的情形中，放松这个双态要求现在就会是可以得到辩护的了。我们理解该要求在这个情形中得不到满足的那个特殊理由。我们能够主张说，因为该情形的这个特殊特征，该要求在此无需得到满足。如果某人的余生会值得过的话，我们能够把拯救他的生命算作有益于他的一个特殊情形。而且如果我们能够在拯救生命的情况下放松这个要求，为什么在给予生命的情况下我们不能做同样的事？如果某人的生活值得过，为什么我们不能把导致他生活算作有益于他的一种特殊情形？

诘难者们或许现在转向

完全比较要求：只有在我们做那对于某人将会更好的事的时候，我们才有益于他。

他们会说："我们导致某人存在并不是在做那对于他将会更好的事。假如我们并未使他存在的话，这本来就对于他不会更糟。"与双态要求的那个最后形式不同的是，这个新要求允许拯救某人的生命能够有益于这个人这一点。我们能够主张，如果某人死亡，即便这并不使他境况更糟，那么这对于他也能够是更糟的。（此外我们会是在拒斥那个卢克莱修式的观点，即只有当一个事件使某人以后受苦或者至少颇有遗憾的时候对他才能够是糟的。）

因为它涵盖拯救生命，这个完全比较要求比双态要求的那个较强形式更令人信服。但是如果我们在那两个特殊情况下都能够放松后者，那么在给予生命的情况下放松前者可能是可以得到辩护的。我们

能够承认，在任一其他种类的情形中，只有当我们做那对于某人将更好的事的时候，我们才有益于他。[3]在给予某人生命的情况下，我们理解为什么那个选择方案对于他来说本来就不会更糟的那个特殊理由。我们可能主张，在这个特殊情形中，该要求无需得到满足。假设我们已经允许拯救某人的生命能够有益于这个人这一点。如果我自己的生命值得过，那么我的生命开始之后在任何一个时刻得到拯救都会是有益于我的。我必须主张说，我的生命恰好在开始了之后得到拯救有益于我，而已经使之开始这一点并不有益于我吗？我能够可辩护地否定这个主张。

导致某人存在之所以是一个特殊情形，是因为备选情形对这个人而言原本不会更糟。我们可能承认，因为这个理由，导致某人存在对这个人而言并不能够更好。但是对这个人而言可能是好的。[4]在这个从“更好”到“好的”的转变中，我们承认那个完全比较要求并未得到满足。但是我们仍然能够做出两类比较。如果某人被导致生存对他而言能够是好的，那么这对这个人而言是如何好法则取决于他的生活如何好——他的生命在多大程度上值得过。而且我们能够作人际间的比较。假设杰克的生活值得过，但幅度并不大。如果他的忧郁症的发作变得越来越频繁和严重的话，那么他就会对他自己的生活是否值得继续下去感到怀疑。相反，吉尔的生活相当值得过。那么我们能够主张说，当我们使杰克存在的时候，这对于杰克是好的，但是这之于杰克却比使吉尔存在之于吉尔少好许多。

这些主张避开了一个常见的诘难。当我们主张说某人被导致存在对于他而言是好的时，我们并不蕴含，假如他没有被造成存在的话，这对于他本来会是糟糕的。而且我们的主张只适用于那些现实中的或者会是现实中的人们。我们并不对那些总是保持纯粹可能的人们做出什么主张。我们并不是在主张说，如果那些可能的人们并不变成现实的人们的话，对于他们而言是糟糕的。

我以这些话来结束。我已经考虑过三件事：从未存在、开始存在和不再存在。我已经提出，就这些而言，开始存在应当与不再存在归为一类。与从未存在不同的是，开始存在和不再存在两者都发生在现实的人们身上。这就是我们为什么会主张对于这些人们而言它们是好的还是坏的的原委。与之对立的主张是说，开始存在应当与从不存在归到一起，而且其中没有哪一个对于人们而言会是好的还是坏的。有时给出的理由是，如果我们本来没有开始存在的话，我们原本根本不会存在，而这对我们而言原本不会更糟。但我们不是在主张说，当它并不发生的时候，开始存在对于人们而言就能够是好的还是坏的。我们的主张是关于当它发生的时候的开始存在。我们承认开始存在与不再存在之间的一个不同。就几乎全部事件而言，如果它们的发生对于人们会是好的，那么对于这些人们而言它们的非发生本来就会是更糟的。但是，我们可以提出，有一个特殊事件，其非发生对于这个实际上的人而言原本就不会是更糟的。这个事件毫不意外地就是这个人将成为实际上的人。

这些论述并不是结论性的。可能有进一步的一些诘难被提出来。我的主张只是说，如果我们认为导致存在这一点能够有益，那么这个信念就是可以得到辩护的。我已经诉诸的有三点。首先，我们无需主张说，如果那些可能的人们从未变成实际上的人的话，那么对于他们而言是糟糕的。正如内格尔所写的："我们全都……已经幸运降生人世。但是……不能够说，没有降生人世是不幸。"[5]其次，如果已经使我的生命恰好在它开始之后得到拯救这一点已经有益于我，那么我并非被迫要否定已经使之开始这一点已经有益于我。从我现在的观点来看，在这两者之间并无深刻的区分。（使我的生命得到拯救已经有益于我这一点可能遭到否定。但是，如果这是所主张的，那么导致某人存在是否能够有益于这个人就变成不相关的。我应该拯救一个溺水的孩子的生命。如果我并非从而有益于这个孩子，那么道德的这个部分就不能够以对人产生影响的术语来解释。）其三，导致某人存在显然

是一个特殊情形。一些人论辩说，这不是什么益处，因为它缺乏所有别的益处都具有的某个特征。但是这个论证乞求于该问题。既然这是一个特殊情形，对任何一个一般规则而言它都可能是一个例外。诉诸某个一般规则恰是假定不能够有什么例外。

还有一个相似的辩论，即关于存在是一个谓词（predicate）还是诸对象可能拥有的一个真正属性。一些人主张，因为它缺乏别的那些谓词所拥有的一些特征，所以存在不是一个谓词。另一些人主张，这只是表明存在是一个特有的谓词。相似地，我们能够主张，导致将会享有值得过的一种生活的某人存在这一点给予这个人一种特有的益处。

---

[1]Williams（2），第85—86页。

[2]在Williams（2），第87页，威廉斯还写道："这里边没有哪一个——包括有计划的自杀想法——需要我对一个其中我永不出现的世界加以反映。就'可能世界'而言……一个人会按照当前的考虑、从他自身的角度有理由宁选其中他活得更长一些而不是较不那么长一些的一个可能世界，或者——像自杀——作出相反的选择；但是他不会有什么宁选一个其中他根本没出现的世界的那种理由。有关他从世界上完全缺位的思想得是不同的一种，对于他在场或者缺位的那个世界之价值的非人称的反映……尽管他能够自我中心地想及他生活得更长一些或者不那么长一些会是什么样子，但是他不能够自我中心地想及他本来根本不存在对于他会是什么。"威廉斯已经指出，如果某人在考虑自杀，他无需比较什么摆在他前面与死了会像什么。这个人只能够决定他是否想要经历所摆在他前面的。这可以被看作是这样一个决定，即对他而言他生活的这个部分是聊胜于无还是糟得无以复加。在刚刚引述的这段话中，威廉斯提到，不能够就一个人的整个生活作出类似的决定或者判断。所提及的理由是，某人"不能够自我中心地想及他本来根本不存在对于他会是什么"。某人能够非常清楚地理解他本来或许永远不存在这样一个可能性。如果有某种他不能够想象的东西的话，它只能够会是像本来永远不存在那样。但是，如果某人断定他的生活一直是值得过的，或宁愿他自己永远没有出生过，那么他则无需作出这个比较——就像那个考虑自杀的人无需比较他的余生与死了会是什么样子。

[3]根据我们对于"益处"（benefit）一词的通常用法这并不为真；但是正如我在第25节中所论证的，按照道德意义上重要的用法这为真。

[4]我把这个提法以及第四编中的许多都归功于J.McMahan。

[5]Nagel（4），第7页。

H 罗尔斯式的诸原则

很多人把罗尔斯只在他的更为广阔的理论语境以内应用的各种原则应用于一些特定的事例。这就是我为什么把这些原则只称为罗尔斯式的（Rawlsian）原则而非罗尔斯的（Rawls's）原则的原委。［马克思曾抱怨他的一些追随者乱用马克思之名（*plus Marxiste que moi*）。罗尔斯本可发出相同的抱怨。］

考虑一下任何一个这样的情形，其中同一些人会存在于不同的结果之中。根据极小中的极大值原则最好的那个结果是其中境况最糟的人们境况最好的结果。我们可能认为，这个结果必定是对于那些境况最糟的人们来说的最好的东西。但并非如此。

假设我是一个医生。取决于我做什么，结果或许是：

（1）杰克、比尔和约翰完全瘫痪。

（2）杰克和约翰被治愈。比尔完全瘫痪。

（3）杰克和比尔被治愈。约翰部分瘫痪。

约翰会比比尔更难治愈，比尔会比杰克更难治愈。这就是为什么我不能够比在结果（2）和（3）中做得更多来治愈这些人的原委。如果我治愈约翰，我就不能够治愈比尔，而且如果我治愈比尔，我只能够部分治愈约翰。

（3）是境况最糟的人境况最好的那种结果。但这并不是对这个人而言最好的那种结果。（3）对于这个人来说比（2）更糟。极小中的极大值原则正确地选择（3）作为最好的结果。但是我们不能够主张说，如果我们遵循极小中的极大值原则，这对于那些境况最糟的人们而言必定是最好的。遵循极小中的极大值原则对于这些人们而言可

能比我们原本做点别的事要更糟。当我们看清这一点的时候，我们仍然能够采取极小中的极大值原则。但是我们必须修正另外两个罗尔斯式的原则。

首先考虑一下差别原则。在结果（3）中有不平等。按照差别原则，造成这样的不平等是不公平的，除非它所产生的结果对于那些境况最糟的人们而言是最好的。在结果（3）中境况最糟的是约翰。在产生这个不平等的时候，我并未产生出对约翰而言最好的那个结果，因为（2）对于他而言更好一些。按照差别原则，结果（3）中的不平等是不公平的。因为相似的理由，结果（2）中的不平等是不公平的。为了避免不公，我必须以两种方式之一行动。我必须要么哪一个都不治愈，要么部分治愈所有三个人。尽管我本来能够完全治愈其中的两个人，部分地治愈第三个人，但是我不应该这么做。这显然是个错误的结论。如果我渴望把它应用于这样的一些特定事例，我们就必须修正差别原则，以便它不再蕴含这样的一些结论。

我无需讨论这一修正，因为差别原则并不适用于我所想象出的结果A和A＋。这个原则只适用于那种既是故意地创造出来的又是可以避免的不平等。A＋中的不平等并非属于此类。

接下来考虑一下我所说的选择原则。这个原则主张说，最好的结果是对于那些境况最糟的人们而言最好的那个结果。在我所举的那个医疗例子中，（3）不能够是最好的结果，因为它对于在（3）中境况最糟的那个人而言比（2）更糟。这显然是个错误的结论。我们必须修正选择原则，以便它不再蕴含这样的一些结论。

显而易见的修正是极小中的极大值原则。这个原则与选择原则之不同恰是在正确地蕴含最好的结果是（3）这一点所需要的那个方面。

本附录所余部分是在我们讨论过前述诸点之后由约翰·布鲁姆执笔所写。

在罗尔斯的著作中，正义原则之一是差别原则，对此罗尔斯这样详细说明：

> 社会和经济的不平等应这样安排，使它们……适合于最少受惠者的利益。

这个公式化陈述并未确切表达出罗尔斯的倾向。假设印度在1800年的时候本来能够具有3部宪法中的任何一部。对别的那些主要利益，每个都会平等地予以分配，只是它们会以下述方式分配经济的和社会的福利：

> 按照宪法（1）印度人和英国人两者都得到100
> 按照宪法（2）印度人得到120，英国人得到110
> 按照宪法（3）印度人得到115，英国人得到140。

显然罗尔斯的差别原则用意在于促进宪法（3）。他有时把差别原则描述为"极小中的极大值规则"，而宪法（3）满足这个规则。但是它并不满足我上文所引的那个陈述。在宪法（3）之下的最少受惠者是印度人，而在宪法（3）之下的那些更大的不平等，与宪法（2）比较，并不适合于印度人的利益。在宪法（2）之下他们的境况原本会更好。

当然，本可以轻易改变措辞以便把罗尔斯的意思更加精确地陈述出来。但是罗尔斯对差别原则的一些论证实际上有赖于他自己的措辞。尽管他的那个主要论证——他那个有关人们在原初状态下会选择极小中的极大规则的主张——并不为真，但是他的其他许多论证都把差别原则促进那些适合于最少受惠者的最大利益的社会安排作为理所当然的事。这些论证在像我所举的那些例子中是失败的。例如，在第

103页上罗尔斯写道：“B”——那个最不利的代表人——“能接受A的较好状况，因为A的利益是通过改善B的前景的方式获得的”。然而英国人的代表人在宪法（3）之下的利益是通过改善印度人的代表人的前景的方式获得的这一点并不为真。

罗尔斯可能并不满意我处理我的例子的方式。他说备选的那些社会安排要以群体状况的标准来比较，或者以代表群体的那些代表人的状况的标准来比较，而不是以个人状况的标准来比较。而且群体仅以其收入和财富水平来界定。（但是请参见第29页，在那里罗尔斯的确允许群体有时以种族来认定。）所以我们应该比较的不是印度人在宪法（2）和（3）之下的地位，而是在这两种情况下由不同的人们构成的最不利群体的地位。在宪法（3）之下的最不利群体的境况好于宪法（2）之下的最不利群体的境况。这就是罗尔斯可能喜欢我们所作的比较。但我的要点是，他的那些论证时常暗中假定我们是在比较同一个体可能占据的那些备选地位。在我所引述的那个论证中，罗尔斯比较的是最不利的代表人的那些备选地位，但是在宪法（2）和（3）这两者之下都没有什么人代表最不利的群体。我们只能够把最不利的代表认作印度人的代表，但是正如我所说的，该论证对他来说并不成功。

显而易见，要证明择取宪法（3）而非（2）这一点为正当，我们得比较两个群体的利益。我们不得不说，择取宪法（3）对于印度人的损失要小于择取宪法（2）对于英国人的损失。所以我们不得不在相当程度上要像功利主义者们那样，以及在相当程度上以差别原则应当避免的方式（第175—183页），权衡不同的人们的利益。如果宪法（3）真的比宪法（2）更好，那么应当说，也只有按照那些像功利主义那样的根据才能够如此。而且如果功利主义的那些根据走向相反方向——如果，比方说，印度人大大多于英国人——我会发现非常难以相信宪法（3）真的更好一些。

约翰·布鲁姆。（若能承蒙罗尔斯发表意见，将不胜感谢。）

I 什么使一个人的生活过得最好

对某人而言什么会是最好的，或什么会是最符合这个人的利益的，抑或什么会使这个人的生活对其而言进行得尽可能地好？对这个问题的那些回答我称之为有关自我利益的诸理论。有三种理论。按照那些快乐论，对某人而言会最好的事是会使其生活最为幸福的事。按照那些愿望实现论，对某人而言会最好的事是会最好地实现其一生中的那些愿望的事。按照目标列表理论，某些事物对我们而言要么是好的要么是坏的，无论我们是否想要这些好的事物或避免这些坏的事物。

狭义快乐主义者虚假地假定，快乐和痛苦是两种不同种类的经验。比较一下满足干渴或者愿望的快乐、听音乐的快乐、解决知识难题的快乐、阅读悲剧的快乐以及知道自己的子女幸福的快乐。这些五花八门的经验并不包含任何独到的共同属性。

痛苦与快乐的共同之处是它们与我们的愿望之间的关系。按照具有合理性上的和道德上的重要意义的"痛苦"一词的用法，所有的痛苦当被经历到的时候全都是人所不欲的，而且一个痛苦越为人所不欲，它就越糟糕或者越大。同样，所有的快乐当被经历到的时候全都是人之所欲求的，它们越是人之所欲求，就越是更好或者更大。这些是偏爱快乐论的主张。按照这个观点，如果两个经验之一得到偏爱，它就更令人愉快。

这个理论无需遵循"痛苦"和"快乐"这些词的通常用法。假设我可以参加一个聚会，享受诸如吃、喝、笑、跳和与朋友叙谈的快乐。我也可以代之待在家里阅读《李尔王》。因为知道这两种选择会是什么样子，我偏爱阅读《李尔王》。说这会带给我更多的快乐便展开了通常的用法。但是按照偏爱快乐论，如果我们添加下面的一些进一步的假定，阅读《李尔王》就会带给我一个更好的夜晚。格里芬引述了一

个更加极端的事例。在弗洛伊德的生命行将结束的时候，他拒绝使用镇痛药物，宁愿在痛苦折磨中思考，而不愿处于头脑混乱的欣快状态。在这两种精神状态中，欣快状态更加令人愉快。但是按照偏爱快乐论，在痛苦折磨之中思考对弗洛伊德则是一个更好的精神状态。在此不要滥用"愉快"一词的意义这一点更加清楚。一位偏爱快乐论者只应当主张，假如弗洛伊德宁愿在痛苦折磨之中清醒地思考，那么如果他的生活像他所偏爱的那样进行则进行得更好一些。[1]

下面考虑一下愿望实现论。其中最简单的形式是非限定论。此理论主张，对某人最好的事是那个会最好地实现一生当中他的所有愿望的事。假设我遇到一个陌生人，他身患一种据认为会致命的疾病。我的同情心被激发起来，而且我强烈地想要这个陌生人获得治愈。我们再未相见过。后来，这个人得到治愈，但我并不知道。按照非限定的愿望实现论，这个事件对我而言是好的，而且使我的生活进行得更好。但这是令人信服的。我们应当拒斥这个理论。

另一个理论只是诉诸有关我们自己生活的那些愿望。我称之为成功论。这个理论只是在一个方面不同于偏爱快乐论。成功论诉诸我们有关自己生活的所有偏爱。一个偏爱快乐论者只诉诸我们生活中内省地可觉察到的那些特征。假设我强烈地不想要被人欺骗。按照偏爱快乐论，如果我相信我不是在受骗对我而言将更好。如果我的信念是虚假的，那将是不相关的，因为这并不给我心的状态造成什么差异。按照成功论，如果我的信念是虚假的，那么对我则更糟。对我自己的生活我有一个强烈的愿望——我不应当这样受到欺骗。如果这个愿望得不到实现，即便我虚假地认为是这样，那对我也是糟糕的。

当这个理论只是诉诸有关我们自己生活的那些愿望的时候，这排除什么可能并不清楚。假设我想要我的生活是这样的，以至于我的所有愿望，无论它们的目标是什么，都得到实现。这可能看起来使应用于我的时候的成功论与非限定的愿望实现论相契合。但是一个成功论者应当主张，这个愿望并非真的有关我自己的生活。这就像某个对象

中的一个真正的变化与所谓的剑桥变化（Cambridge-change）之间的区分。如果在那些能够就一个对象所作的真的陈述中有任何变化，那么这个对象就经历一个剑桥变化。假设当我刮脸的时候伤到了面颊。这在我身上造成了一种真正的变化。它还在孔夫子身上造成一个变化。就孔夫子的情形而言，他生活在一个行星上、后来上面又有一个面颊被割破这一点为真。这纯粹是一个剑桥变化。

假设我是一个流放者，而且不能够与我的子女交流。我想要他们的生活过得好。我可能主张，我想要过其子女生活过得好的某个人的生活。一个成功论者应当再次主张，这并不是真正有关我自己生活的一个愿望。如果我的一个子女被雪崩夺去生命，我并不知情，这对我而言并不糟糕，而且并不使我的生活过得更糟。

一个成功论者本会算计一些相似的愿望。假设我力图带给子女一个好的生活开端。我力求给予他们正确的教育、好的习惯和毅力。再者，我现在是一个流放者，而且我将永远不能够知道我的子女身上发生了什么。假设我的子女的生活过得糟糕，但我并不知道。一个子女发现我给予他的教育使他不能就业，另一个精神崩溃，还有一个沦为小偷。如果我的子女们的生活在这些方面出现失败，而且这些失败部分是我为人父母所犯下的错误的结果，而且我的子女们生活中的这些失败按照成功论会被判定为对我而言是糟糕的。我的那些最为强烈的愿望之一是成功地为人父母。现在发生在我的子女们身上的事情，尽管不为我所知，但是表明这个愿望没有得到实现。我的生活在我最想要它取得成功的方面失败了。尽管我的确并不知道这个事实，对我而言也是糟糕的，而且使这样的一点为真，就是我已经有了一个更糟的生活。这就像我强烈地不想受到欺骗的那个事例一样。即便我永远不知道，如果我既受到欺骗又最终表明是一位不成功的父母，那对我也是糟糕的。这些不是我的意识生活中的那些内省地可以觉察到的差异；所以，按照偏爱快乐论，这些事件对我而言并不是糟糕的。但是按照成功论，它们则是。

接着考虑一下一些人具有的关于他们死后发生什么的愿望。对于一个偏爱快乐论者而言，一旦我死亡，没有什么糟糕的事情能够发生在我身上。一个成功论者应当否定这一点。回到因为我作为父母所犯的错误，我的子女全都过着悲惨生活的那个事例。假设我的那些子女的生活只是在我死后才过得糟糕。最后表明我的生活在我最关心的那个方面是一种失败。一个成功论者在此也应当主张说，这使我过了一个更糟的生活这一点为真。

一些成功论者会拒斥这个主张，因为他们告诉我们要忽略死者的愿望。但是假设我被问到"你想要你即便在死后也是一个成功的父母这一点为真吗？"我会回答"是的"。我的愿望是在我死前还是死后得到实现并不相干。这些成功论者把倘若我的企图遭受失败算作对于我而言是糟糕的，即便我因为是一个流放者，永远不知道这一点。那么当我的企图失败的时候我是否死亡会有什么重要？我的死亡所做的一切不过是确保我将永远不知道这一点。如果我们认为关于我的愿望的非实现我永远不知道这一点并不相关，那么我们就不能够可辩护地主张说，我的死亡造成一种差异。

我现在转向那些因偏爱快乐论和成功论所引发的那些问题与诘难。

我们只应当诉诸某人实际上所具有的那些愿望和偏爱吗？回到我在参加一个聚会和待在家中读《李尔王》之间的选择。假设，明知两种选择会是什么样子，我选择呆在家中。而且假设我后来永不后悔这个选择。按照一个理论，这表明待在家中读《李尔王》带给我一个更好的夜晚。这是一个错误。假如我选择参加晚会的话，我原本永远不会后悔那个选择这一点或许为真。根据这个理论，这本来会表明参加那个晚会给我带来一个更好的夜晚。于是这个理论蕴含着，各种选择本来都比另一个更好一些。既然这个理论蕴含着这样的矛盾，它则必须加以修正。显而易见的修正是，不仅诉诸我所选择的方案中的我的

实际上的偏爱，而且诉诸假如我作出了别的选择的话本来会有的偏爱。[2]

在这个例子中，无论我选择哪一个方案，我永远不会后悔这个选择。如果这是真的，我们仍然能够主张那些选择之一会给予我一个更好的夜晚吗？按照某些理论，当在这两个备选中我会有这样的一些矛盾偏爱的时候，没有哪一个对于我而言更好或者更糟。当我的那些矛盾的偏爱之一本来会更强烈的时候，这一点就并不令人信服。假设，如果我选择参加晚会，我将只是适度地高兴我作出了这个选择，但是，如果我选择待在家中看《李尔王》，我将非常高兴。如果这是真的，读《李尔王》就带给我一个更好的夜晚。

无论是诉诸偏爱快乐论还是成功论，我们都不应当只是诉诸我们实际上所具有的那些愿望或者偏爱。我们还应当诉诸在不同的时刻向我敞开的各种各样的备选之中我本来会有的愿望和偏爱。如果这些备选中的一个是我的那些最为强烈的愿望和偏爱得到实现的，那么它对我而言就是最好的那个备选。这允许我们主张说，某个备选生活本来对于我而言会更好，纵使在我的整个实际生活中，我高兴的是我选择了这个生活而不是这个备选生活。

既适用于偏爱快乐论又适用于成功论的区分还有另一个。如果这些理论诉诸某人的所有愿望，无论是实际上的还是假设的愿望，也无论是关于他的心的状态的还是关于他的生活的，它们都是累积性的。在决定哪一个备选会产生愿望实现的最大净总额的时候，我们分配给各个得到实现的愿望某个正数，各个并未得到实现的愿望某个负数。这些数值有多大取决于所谈论的问题中的那些愿望的强度。（就成功论的情况而言，它诉诸过去的愿望，可能还取决于拥有这些愿望有多长。正如我在第八章中所提的，这可能是这一理论的弱点。对偏爱快乐论而言，该问题并不出现，因为它只诉诸我们在不同的时候有关我们当前心的状态所拥有的那些愿望。）愿望实现总净额就是那些正数

减掉那些负数后的总额。如果我们能够比较不同愿望的相对强度，那么这个计算在理论上说是可以进行的。对数值而言单位的选择对结果并不造成什么差异。

这两个理论的另一个版本并不以这种方式诉诸一个人有关他自己的生活所具有的所有愿望和偏爱。它只是诉诸全局性的而不是局部性的愿望和偏爱。如果一个偏爱是关于一个人的被当作整体的某个生活部分，或者是关于一个人的整个生活，那么它就是全局的。我认为，这些理论的那些全局性的版本是更加令人信服的。

考虑一下这样一个例子。明知你接受一个累积性的理论，我告诉你我准备使你的生活过得更好。我将给你注射一种上瘾药物。从现在开始，每天早晨你都将带着一种想要注射另一针这种药物的强烈愿望醒来。有这个愿望这一点本身将既不是愉快的也不是痛苦的，但是如果该愿望不能够在1个小时之内得到实现，那么它将变得十分痛苦。这不是什么关心的理由，因为我将把这种药物充裕地供给你。每天早晨，你将能够即刻实现这个愿望。注射及其后作用也会是既不愉快又不痛苦。你将像现在那样度过你所余下的日子。

有关这个事例那些累积性的理论会蕴含什么？我们能够令人信服地假设，你不会欢迎我的提议。你会宁愿不对这种药物变得上瘾，即便我向你保证你将永远不会缺乏供应。我们还能够令人信服地假设，如果我继续下去，你将总是后悔你变得对这种药物上瘾。但很有可能的是，你的不要变得上瘾的最初愿望，以及对你变得上瘾了的后悔，不会像你每天早晨对另一次注射的那些愿望一样强烈。鉴于像我所描述的那样一些事实，你宁愿不要变得上瘾的理由不会是非常强烈的。你可能不喜欢对任何事物上瘾的想法；而且你会后悔总要记着携带足量药物所牵扯到的那个小小的不方便。但是这些愿望或许远远弱于每天早晨你会有的对新注射一针药物的那些愿望。

根据那些累积性的理论，如果我使你成为一个瘾君子，我将是在增加你的愿望实现的总额。我将是在造成你的愿望之一，即你的那个

不要变成瘾君子的愿望，不会得到实现，这个愿望在我的行动之后变成一个有待矫治的愿望。但是我也将是在给予你无限的、极端强烈的一个愿望系列，每天早晨一个愿望，所有这些愿望你都可以实现。所有这些愿望的实现会压倒你的那些不要变成瘾君子、要加以矫治的愿望的非实现。按照那些累积性的理论，通过把你变成一个瘾君子，我将是在有益于你——是在使你的生活过得更好。

这个结论并不令人信服。具有这些愿望，以及使它们得到实现，既不是愉快的，也不是痛苦的。我们无需是些快乐主义者就可以更加令人信服地相信，对你而言具有和实现这一强烈愿望系列绝非更好一些。

那些累积性的理论可以得到修正，以便满足这个诘难吗？那些上瘾性的愿望有某个特征会证明当我们计算你的愿望实现总额的时候我们应当忽略它们这一主张是正当的吗？我们可能主张说，它们之所以能够被忽略是因为它们是些你宁愿不要具有的愿望。但这并非一个可以接受的修正。假设你处于巨大的痛苦之中。你现在具有一个不要处于现在所处状态的强烈愿望。按照我们修正过的理论，如果你宁愿没有这个愿望，一个愿望就不作数。这必定适用于你的那个不要处于所处状态的强烈愿望。你会宁愿没有这个愿望。假如你并非不喜欢你所处的状态，它就不会是痛苦的。既然我们修正过的理论并不计算你会宁愿不具有的那些愿望，那么它荒谬地蕴含着对于你而言处于巨大的痛苦之中不能够是糟糕的。

可能有别的修正会满足这些诘难。但是诉诸偏爱快乐论和成功论两者的全局性的版本则更简单一些。这些只是诉诸某人有关他的被视作一个整体的某部分生活的愿望或者关于他的整个生活的愿望。在我使你变成瘾君子的事例中，这些全局性的理论给予我们那个正确的答案。你会宁愿不要变得上瘾，而且你后来会宁愿不再上瘾。这两个全局性的理论所诉诸的偏爱惟有这些。它们忽略你每天早晨对新注射一

针的特定愿望，因为在形成你的全局性偏爱的时候你已经考虑过这些愿望。

这个想象出来的上瘾例子在其本质要素上与别的无数事例相似。下述（1）和（2）两者都为真的事例有无数：（1）如果某人的生活走了两条路中的一条，这会产生局部性愿望实现的一个更大的总额，但是（2）另一个备选是他会全局性地偏爱的，无论他的生活所走的路是哪一条。

我不是描述无数实际情形中的另一个事例，而是提一个想象的事例。这个一生之内的事例类似于我在第四编中所讨论的那个令人讨厌的结论。假设我要么可以享有50年极高质量的生活，要么可以享有勉强值得过的无限年数的生活。在第一个备选情况中，我的那50年按任何理论说会过得极端地好。我会非常幸福，会获得美好的事物，做大量的善事，爱许多人并为许多人所爱。在第二个备选情形之中，我的生活尽管不怎么样，但是总会是值得过的。有关这个生活不会有什么糟糕的，而且它每天会包含一些小快乐。

按照那些累积性的理论，如果第二种生活足够长的话，对我而言就会更好。在这一生之内的每一天中，我有一些关于我的生活的愿望得到实现。在第一种备选情况的50年中，会有非常大的一个局部性的愿望实现总额。但这会是一个有限的总额，而且最终它会被我的无限长的第二个选择方案中的愿望实现总额所压倒。这一点可以更加简单地这样来表达。第一个选择方案情况会是好的。在第二个选择方案情形中，假如我的生活是值得过的，那么额外过的每一天对我而言都是好的。如果我纯粹把对于我而言无论什么好的东西都累加在一起，这些额外天数的某个数目会产生最大的总额。

我并不认为第二个备选会带给我一种更好的生活。我因而拒斥那些累积性的理论。很有可能的是，在这两个选择方案中，我会全局性地偏爱第一个。既然那些全局性的理论会蕴含着第一个选择方案给予我一种更好的生活，那么这些理论在我看来就更加令人信服。[3]

现在转向我所提到的第三种理论：目标列表理论。根据这个理论，对于人们而言某些事物是好的或者是坏的，无论这些人是否会想要具有这些好的事物或者想要避免这些坏的事物。那些好的事物或许包括道德上的善、合理的行动、一个人的能力的发展、生养子女和慈父贤母、知识和对真正的美的意识。那些坏的事情或许包含遭到背叛、操纵、诽谤、欺骗，被剥夺自由或者尊严，以及享受施虐的快乐或者在实为丑陋的东西之中寻找美学上的快感。[4]

一个目标列表理论者可能主张说，他的理论与成功论的那个全局性的版本相契合。按照这个理论，什么会使我的生活过得最好，取决于假如我知道有关备选的相关事实的话，我现在以及在各种各样的备选中会偏爱的东西。一个目标列表理论者可能说，最为相关的事实就是那些刚刚提到的事实——有关对于我而言什么会好或什么会坏的事实。而且他可能主张说，知道了这些事实的任何一个人会想要对他而言好的东西，而且想要避免对他而言坏的东西。

即便这是真的，尽管目标列表理论会与成功论相契合，但是这两个理论仍然会有不同。一个成功论者会拒斥对该契合的这种描述。按照他的理论，无论人们的偏爱是什么，对于他们来说没有什么东西是好的或者坏的。只有当一个人倘若知道了事实，会想要避免某物的时候，这个某物对他而言才是坏的。而那些相关的事实并不包括那些目标列表论者所引述的那些所声称的事实。按照成功论，一个人如果受了骗，而且这不是他所想要的，对这个人来说就是坏的。目标列表理论则作出相反的主张。人们想要不被欺骗，乃因这对他们而言是坏的。

正如这些评论所蕴含的，一方面在偏爱快乐论和成功论之间，另一方面在和目标列表理论之间，存在着一个重要差异。前两种理论对自我利益的解释是纯粹描述性的——并不诉诸有关价值的事实。这个解释只诉诸一个人在充分掌握有关备选情况的那些纯粹非评估性的事

实的时候做什么以及会偏爱什么。相反，目标列表理论则直接诉诸它所宣称的关于价值的那些事实。

在这些理论之间作出选择，我们必须决定给予所设想出来的这样的一些事例多大权重，在这些事例中某人在完全知情的情况下的偏爱是稀奇古怪的。如果我们能够诉诸这些事例，它们就会使人对偏爱快乐论和成功论产生怀疑。考虑一下罗尔斯所想象出来的那个想要把毕生的精力花在数不同草坪上草叶数目的那个人。假设这个人知道倘若他代之钻研应用数学中某个特别有用的部分本会取得巨大进步。尽管他本会取得这样的一些有意义的结果，但是他仍然偏爱去继续数草叶。按照成功论，如果我们允许这个理论涵盖所有想象出的情形，那么假若这个人继续数草叶而不是取得巨大的、有用的数学成就，对于他而言则会更好。

反例或许更加令人不快。假设某人在知道各个备选的情况下最偏爱的是这样一种生活，其中在不被发觉的情况下他给他人造成了他所能够造成的那么大的痛苦。按照成功论，这样的一种生活对这个人而言会是最好的。

我们或许并不能够接受这些结论。因而我们应该放弃这个理论吗？这正是西季威克所做的，但是那些引述他的人们却很少注意到这一点。他提到"如果一个人能够准确地预见并且在当下的想象中充分地表现出所有对于他来说是可能的不同行动选择的所有后果，他的将来的整体的善也就是他现在总的说来会去欲求和追求的东西"。[5]正如他所评论的："这样获得的'善'的概念有一种理想的成分，它是某种并不是始终被人们实际欲求和追求的东西。但是，这种理想的成分完全可以借助于实际的或假设的事实来解释，并且不引入任何与有关存在的判断有根本区别的价值判断。"西季威克然后拒斥这个说法，他主张说，对于某人而言的终极的善就是假如这个人的愿望与理性相协调的话他会欲求的东西。西季威克认为，这个短语是需要排除其中有的人的愿望是非理性的那些情形。他认为，有一些事物，我们有好理由去

欲求；另有一些事物，我们则有好理由不去欲求。这些或许就是诸目标列表理论所认为的那些对于我们而言或者好或者坏的事物。

假设我们同意，在一些所设想的情形中，在知道各种选择方案的情况下，某人现在和后来最想要的不会是对于他来说最好的东西。如果我们接受这个结论，可能看起来我们必须既要拒斥偏爱快乐论又要拒斥成功论。大概就像西季威克一样，我们必须对我们能够合理地欲求的东西有所限制。

或许代之有人主张说，我们能够拒斥对这样的一些情形的诉诸。也许有人主张说，假如人们知道相关事实，那么他们事实上会偏爱的东西总归是那些我们会作为对于他们而言真正好的东西加以接受的某种东西。这是一个好的回应吗？如果我们同意，在那些想象出的情形中，某人会偏爱的东西对于他而言或许是坏的某种东西的话，那么在这些情形中我们就已经放弃我们的理论了。如果是这样，我们能够凭借说出在那些实际情形中它不会误入歧途这一点来为我们的理论辩护吗？我认为这并非是一个充分的辩护。但是对于这个问题我在此将不予探索。

这个诘难可能适用于偏爱快乐论，但是力量要小一些。按照这个理论，对于某人而言能够是好或是坏的东西惟有他的意识生活的那些可以察觉的特征。这些就是他其时要么想要要么不想要的那些特征。我在上文中问及，是因为他们宁愿不被欺骗所以对于他们而言被欺骗是坏的，还是因为对于他们而言是坏的所以他们宁愿不被欺骗？考虑一下那个可以比较的有关痛苦的问题。一些人业已主张说，痛苦是内在地坏的，而且这就是我们不喜欢它的原委。正如我所提到的，我对这个主张持有怀疑。在服用某些种类的药物之后，人们宣称他们的感觉质量并未改变，但是他们不再不喜欢这些感觉。我们会把这样的一些药物当作有效的止痛剂。这提示着，痛苦的坏性寓于其不被喜欢之中，而且并非因为它是坏的才不被喜欢。这些观点之间的歧见会需要更多的讨论。但是，如果第二个观点更好，那么下述主张则更加令人

信服，就是说，无论一个人想要或者不想要经历的是什么——无论我们发现他的那些愿望如何古怪——都应当算作对于这个人而言是真正快乐的或痛苦的，而且都应当算作对于他而言因为那个理由是好的或是坏的。（可能仍然有一些情形，其中下述主张是令人信服的，就是说，假如某人享受某些种类的经验的话，对他而言会是坏的；例如，关于施虐狂式的快乐或许就可以做此主张。但是这样的情形可能少之又少。）

相反如果我们诉诸成功论，那么我们就不仅关心经验到的我们的意识生活的质量。我们还关心这样的一些事情，诸如我们是否在取得我们努力取得的东西，我们是否在受到欺骗，等等。当考虑这个理论的时候，我们能够更加经常令人信服地主张，即使某人知道相关事实，他的那些偏爱也可能误入歧途，不能与对他而言会是好的或者会是坏的东西相呼应。

我们应当接受这些不同理论之中的哪一个？在此我不想尝试给出答案。但是我将提及另一个理论，以此作为结束。这个理论或许可以被当作结合了这些相互冲突的理论中最令人信服的东西。那些对这个问题发表过见解的人们的分歧如此之大，这是一个惊人的事实。很多哲学家已经成为令人信服的快乐主义者；在相同程度上，还另有许多哲学家深信快乐论是一个粗劣的错误。

一些快乐主义者已经如下这般地达到他们的观点。他们考虑一个相对立的观点，诸如这样一个主张，就是说，对于某人而言好的东西是具有知识、参与合理的活动、鉴赏真正的美。这些快乐主义者问："假如心的这些状态并未带来什么享受，而且假如处于心的这些状态中的那个人丝毫并不欲求它们继续，那么心的这些状态会是好的吗？"既然他们的答案是否定的，那么他们得出结论，心的这些状态的价值必定寓于它们之被喜欢之中，以及寓于它们唤起对它们之继续的愿望之中。

这个推理假定，一个整体的价值就是他的诸部分的价值的总和。如果我们去除该种快乐主义者所诉诸的那个部分，所剩下的看起来就没了什么价值，因此快乐论是真理。

相反假设，这一点更加令人信服，就是一个整体的价值可能并不纯粹是它的诸部分价值的总和。那么我们或许主张，对人们而言最好的东西是一种混成物。它并不只是人们处于他们想要处于的那些意识状态之中。它也不是人们具有知识、参与合理的活动、鉴赏真正的美等等。对于某人来说好的东西既不仅仅是那些快乐主义者所主张的东西，也不仅仅是那些目标列表理论者所主张的东西。我们可能认为，假如我们有的是这两者中的任何一个，而没有另一个的话，那么我们所拥有的就几乎没有或者根本没有价值了。例如，我们可能主张，对于某人而言或好或坏的东西是拥有知识、参与合理的活动、经历互爱和欣赏真正的美，而同时强烈地想要的就是这些东西。按照这个观点，这个争议中的各方只是看到了真理的一半。各自把必要的东西当作充分的东西提了出来。带有其他许多种类的对象的快乐没有什么价值。而且如果知识、合理行动、爱或对美的意识缺乏快乐，那么它们之中就没有什么价值。有价值的东西或者对于某人而言好的东西要具有这两者；从事这些活动，而且强烈地想要如此从事活动。[6]

---

[1]Griffin（1）。

[2]参见P·布里克《谨慎》一文，载《哲学杂志》，1980年7月。

[3]再参见布里克。

[4]例如，参见Moore and Ross（2）。

[5]Sidgwick（1），第111—112页。

[6]参见Edwards通篇。柏拉图在《斐利布篇》中有相似的提法。至于那些有关自我利益的各种不同理论的更深刻的讨论参见J·格里芬《幸福安康》（牛津大学出版社，1986年）。

J 佛家的观点[1]

谈话伊始，国王[2]就非常委婉地询问比丘的名字，得到下面的回答："人们知道我是'那先'；我的宗教生活中的同道称我为'那先'。尽管我的父母给我取名'那先'……但是它只是一个称呼，一个话语形式，一个描述，一个惯用法。'那先'只是一个名，因为此处并不见人。"[3]

> 魔啊，你以为有情确实存在？
> 其实你受到了错误观念的误导。
> 这束元素空我，
> 其中并无有情，
> 恰如一组木件，
> 得到其名曰车，
> 故吾给予元素，
> 名以空幻存在。[4]

佛陀如此说："兄弟啊，行动确实存在，其后果亦然，但是行动之人并不存在。没有人抛弃这组元素，没有人设想它们的一个新组合。不存在什么个体，有的只是给予一组元素的一个惯用名。"[5]

世亲（Vasubandhu）：……当佛陀说"我自己就是这个夫子Sunetra"的时候，他的意思是说他的过去和现在属于诸无间之有（momentary existences）的同一个谱系；他并不是说前面的元素并未消失。恰如当我们说"已经看到烧掉那事物的这同一把火已经临到此一尘物"的时候，该火并非同一火，而是由于忽视这个差异我们展转称其诸刹那（moments）的连续性为火而已。[6]

犊子部（Vatsiputriya）：如果没有灵魂（Soul，人我），谁在记念？世亲："记念"这个词是什么意思？犊子部：它的意思是凭记忆把

握一个尘物。世亲：这个“凭记忆把握”某物与记忆有什么不同吗？犊子部：它是一个贯穿记忆的行动者。世亲：借此记忆得以产生的动因我们刚才已经解释过。忆念产生之因是一种适当的心境，仅此而已。犊子部：但是当我们使用“Caitra记得”这个表达的时候，是什么意思？世亲：在Caitra这个名所指定的诸法之流中，一个忆念出现。[7]

在佛教用语中个体有一个专门术语，这个意在提示佛教观点与别的理论之间差异的术语就是不共业（santana），亦即一种“预流”。[8]

犊子部：什么是实有，什么是名有？世亲：如果某物（像一个独立的元素）自己存在，那么它就具有一种实有。但是如果某物代表（这样的一些元素的）一种组合，它就是一种名有。[9]

实际只是名与色，
没有人和有情的存在；
空如造作的木偶，
一团苦，如草木。[10]

---

[1]出于译文前后呼应和方便理解的原因，所涉及的佛典基本上按照英文直接译出。——译者

[2]此处指古印度西北舍竭国（大约在当今巴基斯坦西北部白沙瓦一带）国王弥兰陀，他与那先关于佛法的问答，汉译本称《那先比丘经》，南传巴利文称《弥兰陀王问经》。——译者

[3]《转识论》，转引自Collins，第182—183页。

[4]《魔》，转引自T·斯切尔巴茨基，《佛教徒的灵魂论》，载《俄罗斯科学院简报》，1919年，第839页。

[5]《世亲》，转引自斯切尔巴茨基，《佛教徒的灵魂论》，载《俄罗斯科学院简报》，1919年，第845页。

[6]转引自斯切尔巴茨基，《佛教徒的灵魂论》，载《俄罗斯科学院简报》，1919年，第851页。

[7]转引自斯切尔巴茨基，《佛教徒的灵魂论》，载《俄罗斯科学院简报》，1919年，第853页。

[8]参见Collins，第247—261页。
[9]T·斯切尔巴茨基，《佛教的中心概念》，皇家亚洲学会，伦敦，1923年，第26页。
[10]《清净道论》，转引自Collins，第133页。

专业术语汉英对照表

A

安全政策 safe policy

B

标准观点 standard view

不顾威胁者 threat-ignorer

不偏不倚的仁爱 impartial benevolence

部分依从理论 partial compliance theory

C

差别原则 difference principle

常识道德 common-sense morality

成功论 success theory

成就欲 desires for achievement

传递性的 transitive

纯粹加法悖论 mere addition paradox

从不克己的 never self-denying

D

搭便车者 free raider

单独的施虐者 single torturer

当前的自我主义 egoism of the present，缩写EP

当前目标论 present-aim theory

当前偏向 bias towards the present

道德的自我放纵 moral self-indulgence

道德怀疑论 moral scepticism

笛卡儿式的自我 Cartesian ego
第一论证 first argument
电子传输器 teletransporter
动机 motives
对人产生影响的观点 person-affecting view
对行动者中立的 agent-neutral
对自身的偏向 bias in one's own favour
多因素决定 overdetermination
多元主义的，多元论者 pluralist

E

二元论者 dualist

F

反论证 counter-argument
反向变量观 backward variation view
反应理论 reaction theory
非后果论 non consequentialism
非还原论者 non-reductionist
非还原论者的观点 non-reductionist view
非人称的，非人称性的 impersonal
非同一性问题 non-identity problem
非限定的愿望实现论 unrestricted desire-fulfillment theory
非限定论 unrestricted theory
父母两难 parent's dilemma
复制品 replica

G

概率贴现率 probabilistic discount rate
工具论 instrumental theory
工具论的当前目标论 IP

工具性目标 instrumental aim
公共的善 public goods
功利主义 utilitarianism
功利主义者的 utilitarian
贡献者两难 contributor's dilemma
古典自利论 classical self-interest theory

H

合理的仁爱 rational benevolence
后果论 consequentialism
还原论者 reductionist
还原论者的观点 reductionist view

J

机会成本 opportunity costs
极大中的极大值原则 maximax
极端主张 extreme claim
极小中的极大值原则 maximin
集体性的后果论 collective consequentialism
近期偏向 bias towards the near
精英主义观点 elitist view
决定论者 determinist
均量原则 average principle

K

克己 self-denying
空间贴现率 spatial discount rate
快乐论 hedonistic theory
快乐主义者 hedonist

L

老右 righty

老左 lefty
类记忆 quasi-memory
类意图 quasi-intention
累积性的 summative
理想的审慎 ideal deliberation
理想动机理论 ideal motive theory
理想契约主义 ideal contractualism
理想行动理论 ideal act theory
理由—相对性两难 reason-relativity dilemma
联合谱系 combined spectrum
两水准质的观点 two-level quality view
量的同一性 numerically identity
伦理学 ethics
罗尔斯式的原则 Rawlsian principles

M

冒险政策 risky policy
没有价值的水准 valueless level
描述性观点 descriptive view
目标列表理论 objective list theory

P

批判性的当前目标论 critical present-aim theory，缩写CP
偏爱快乐论 preference-hedonism
平等原则 equality principle

Q

起源的观点 origin view
囚犯的两难困境 prisoner's dilemma
取代 replacement

R

人称的，人称性的 personal
人际两难 interpersonal dilemma

S

善行原则 principle of beneficence
上班族的两难困境 commuter's dilemma
社会贴现率 social discount rate
审慎论 deliberative theory
审慎论的当前目标论 DP
时际两难 intertemporal dilemma
时间上中立的 time-neutral
时间贴现率 temporal discount rate
实践动机理论 practical motive theory
实践行动理论 practical act theory
实施威胁者 threat-fulfiller

T

T设目标 T-given aims
统一理论 unified theory

W

完全依从理论 full compliance theory
唯物主义 materialism
唯心论者 idealist
未来偏向 bias towards the future
温和主张 moderate claim
无差异观点 no-difference view
无害的施虐者 harmless torturer
物理谱系 physical spectrum
物理主义 physicalism
物理准则 physical criterion

X

狭义的善行原则 narrow principle of beneficence

相对于时间的 time-relative

相对于行动者的 agent-relative

相继的自我 successive selves

协同问题 co-ordination problems

心理决定论 psychological determinism

心理谱系 psychological spectrum

心理准则 psychological criterion

心理自我主义 psychological egoism

行动的理由 reasons for acting

行善者 do-gooder

行为功利主义 act utilitarianism

修正的自利论 revised self-interest theory

选择原则 selection principle

Y

意向 disposition

愿望 desire

元伦理学 meta-ethics

愿望实现论 desire-fulfillment theory

Z

支线事例 branch-line case

质的同一性 qualitatively identity

终极性目标 ultimate aim

主线 main-line

自败的，自我挫败的 self-defeating

自利，自我利益 self-interest

自利论 self-interest theory，缩写S

自明之理 truism
自谦的 self-effacing
自由意志 free will
总量原则 total principle
总体份额 share-of-the-total
最近连续者图式 closest continuer schema

参考文献

ADAMS (1): R. M., 'Must God Create the Best?', *Philosophical Review* 81, No. 3, July 1972.

ADAMS (2): R. M., 'Motive Utilitarianism', *The Journal of Philosophy*, 12 Aug. 1976.

ADAMS (3): R. M., 'Existence, Self-Interest, and the Problem of Evil', *Nous* 13, 1979.

ANSCHUTZ : R. P., *The Philosophy of J. S. Mill*, Oxford, Clarendon Press, 1953.

ANSCOMBE (1): G. E. M., *Intention*, Ithaca, N.Y., Cornell University Press, 1957.

ANSCOMBE (2): G. E. M., 'Who is Wronged?', *The Oxford Review*, 1967.

AYER (1): A. J., *The Concept of a Person and Other Essays*, London, Macmillan, 1964.

AYER (2): A. J., *Philosophical Essays*, London, Macmillan, 1965.

BAIER (1): K., *The Moral Point of View*, Ithaca, N.Y., Cornell University Press, 1958.

BAIER (2): K., 'Ethical Egoism and Interpersonal Compatibility', *Philosophical Studies* 24, 1973.

BAIER (3): K., 'Rationality and Morality', *Erkenntnis*, 1977.

BARRY (1): B., *Political Argument*, London, Routledge and Kegan Paul, 1965.

BARRY (2): B., 'Justice Between Generations', in P. M. S. Hacker and J. Raz, eds., *Law, Morality, and Society: Essays in Honour of H. L. A. Hart*, Oxford, Clarendon Press, 1977.

BARRY (3): B., 'Rawls on Average and Total Utility: A Comment', *Philosophical Studies* 31, 1977.

BARRY (4): B., *Sociologists, Economists, and Democracy*, London, 1970.

BAYLES : M. D., ed., *Ethics and Population*, Cambridge, Mass., Schenkman, 1976.

BENDITT : T., 'Happiness', *Philosophical Studies* 25, 1974.

BENNETT : J., *Kant's Analytic*, Cambridge University Press, 1966.

BENTHAM : J., *An Introduction to the Principles of Morals and Legislation*, first published in 1789, reprinted in *Utilitarianism*, M. Warnock, ed., London, Collins, 1962.

BLANSHARD : B., 'Sidgwick The Man', *The Monist*, 1974.

BOGEN : J., 'Identity and Origin', *Analysis* 26, Apr. 1966.

BRAMS : S. J., 'Newcomb's Problem and Prisoners' Dilemma', *Journal of Conflict Resolution* 19, No. 4, Dec. 1975.

BRANDT (1): R. B., ed., *Value and Obligation*, New York, Harcourt, Brace, and World, 1961.

BRANDT (2): R. B., *A Theory of the Good and the Right*, Oxford, Clarendon Press, 1979.

BRAYBROOKE : D., 'The Insoluble Problem of the Social Contract', *Dialogue*, Mar. 1976.

BRENNAN (1): A. A., 'Personal Identity and Personal Survival', *Analysis* 42, Jan. 1982.

BRENNAN (2): A. A., 'Survival', *Synthese*, 1984.

BROAD (1): C. D., 'On the Function of False Hypotheses in Ethics', *Ethics*, 1915.

BROAD (2): C. D., *The Mind and its Place in Nature*, London, Routledge and Kegan Paul, 1949.

BROAD (3): C. D., *Five Types of Ethical Theory*, Littlefield, Adams, 1959.

BROOME (1): J., 'Indefiniteness in Identity', *Analysis*, in or before 1984.

BROOME (2): J., 'Rational Choice and Value in Economics', *Oxford Economic Papers* 30, Nov. 1978.

BUCHANAN : A., 'Revolutionary Motivation and Rationality', *Philosophy and Public Affairs* 9, No. 1, Fall 1979.

BUTLER : J., *The Analogy of Religion*, first Appendix, 1736, reprinted in PERRY (1).

CHISHOLM : R., 'Reply to Strawson's Comments', in H. E. Kiefer and Milton K. Munitz, eds., *Language, Belief, and Metaphysics*, Albany, N.Y., State University of New York Press, 1970.

COLLINS: S., *Selfless Persons*, Cambridge University Press, 1982.
DANIELS: N., 'Moral Theory and the Plasticity of Persons', *The Monist* 62, No. 3., July 1979.
DENNETT: D. C., *Brainstorms*, Bradford Books, 1978.
DESCARTES: R., *Meditations*, translated by E. S. Haldane and G. R. T. Ross, Cambridge University Press, 1969
DUMMETT: M. A. E., 'Wang's Paradox', *Synthese* 30, 1975.
DWORKIN (1): R., 'What is Equality? Part 1: Equality of Welfare', *Philosophy and Public Affairs* 10, No. 3, Summer 1981.
DWORKIN (2): R., 'What is Equality? Part 2: Equality of Resources', *Philosophy and Public Affairs* 10, No. 4, Fall 1981.
EDGLEY: R., *Reason in Theory and Practice*, London, Hutchinson, 1969.
EDIDIN: A., 'Temporal Neutrality and Past Pains', *The Southern Journal of Philosophy* 20, No. 4, Winter 1982.
EDWARDS: R. B., *Pleasures and Pains*, Ithaca, N.Y., Cornell University Press, 1979.
ELLIOT: R., 'How to Travel Faster Than Light?', *Analysis* 41, Jan. 1981.
EVANS (1): G., 'Can There Be Vague Objects?', *Analysis* 38, 1978.
EVANS (2): G., *The Varieties of Reference*, Oxford, Clarendon Press, 1982.
EWING: A. C., 'Suppose Everybody Acted Like Me', *Philosophy* 28, Jan. 1953.
FEINBERG: J., 'The Rights of Animals and Future Generations', in William K. Blackstone, ed., *Philosophy and Environmental Crisis*, Athens, Ga., University of Georgia Press, 1974.
FINDLAY: J., *Values and Intentions*, London, George Allen and Unwin, 1961.
FISHKIN (1): J. S., *The Limits of Obligation*, New Haven, Ct., Yale University Press, 1982.
FISHKIN (2): J. S., 'Justice Between Generations: the Dilemma of Future Interests', *Social Justice* 4, 1982.
FOOT: P., 'Morality as a System of Hypothetical Imperatives', *The Philosophical Review* 81, 1972.
FORBES (1): G., 'Origin and Identity', *Philosophical Studies* 37, 1980.
FORBES (2): G., 'Thisness and Vagueness', *Synthese* 54, 1983.
GALE (1): R. M., *The Language of Time*, London, Routledge and Kegan Paul, 1968.
GALE (2): R. M., ed., *The Philosophy of Time*, London, Macmillan, 1968.
GAUTHIER (1): D., 'Morality and Advantage', *Philosophical Review*, 1967.
GAUTHIER (2): D., *Practical Reasoning*, Oxford, Clarendon Press, 1962.
GAUTHIER (3): D., 'Reason and Maximization', *Canadian Journal of Philosophy* Mar. 1975.
GAUTHIER (4): D., *Morals By Agreement* (provisional title), Oxford, Clarendon Press, forthcoming.
GERT: B., *The Moral Rules*, New York, Harper and Row, 1973.
GLOVER (1): J. C. B., *Causing Death and Saving Lives*, Harmondsworth, Penguin Books, 1977.
GLOVER (2): J. C. B., 'It Makes No Difference Whether Or Not I Do It', *Proceedings of the Aristotelian Society, Suppl. Vol.* 49, 1975.
GLOVER (3): J. C. B., *What Sort of People Should There Be?*, Harmondsworth, Penguin Books, 1984.
GODWIN: W., *Political Justice*, 1793, and Oxford University Press, 1971.
GOODIN: R. E., 'Discounting Discounting', *Journal of Public Policy* 2, Pt. 1, Feb. 1982.
GOSLING: J. C. B., *Pleasure and Desire*, Oxford, Clarendon Press, 1969.
GRICE: G. R., *The Grounds of Moral Judgement*, Cambridge University Press, 1967.
GRIFFIN (1): J. P., 'Are There Incommensurable Values?', *Philosophy and Public Affairs* 7, No. 1, Fall 1977.
GRIFFIN (2): J. P., 'Modern Utilitarianism', *Revue Internationale de Philosophie*, No.

141, 1982.

GRIFFIN (3): J. P., 'A Substantive Theory of Rights', unpublished paper.

GRIFFIN (4): J. P., 'Is Unhappiness Morally More Important Than Happiness?', *Philosophical Quarterly* 29, 1979.

GRIM: P., 'What Won't Escape Sorites Arguments', *Analysis* 42, 1982, p. 38.

GUTTENPLAN: S., ed., *Mind and Language*, Oxford, Clarendon Press, 1975.

HAIGHT: G. S., *George Eliot: A Biography*, Oxford University Press, 1978.

HAKSAR: V., *Equality, Liberty, and Perfectionism*, Oxford, Clarendon Press, 1979.

HARDIN: G., 'The Tragedy of the Commons', *Science* 162, 13 Dec. 1968.

HARE (1): R. M., *Moral Thinking*, Oxford, Clarendon Press, 1981.

HARE (2): R. M., 'Ethical Theory and Utilitarianism', in H. D. Lewis, ed., *Contemporary British Philosophy*, London, George Allen and Unwin, 1976, reprinted in SEN AND WILLIAMS .

HARE (3): R. M., 'Pain and Evil', *Proceedings of the Aristotelian Society, Suppl. Vol.* 38, 1964

HARE (4): R. M., *Freedom and Reason*, Oxford University Press, 1963.

HARE (5): R. M., 'Abortion and the Golden Rule', *Philosophy and Public Affairs* 4, No. 3, Spring 1975.

HARE (6): R. M., *Essays on Philosophical Method*, London, Macmillan, 1971.

HARE (7): R. M., *Practical Inferences*, London, Macmillan, 1971.

HARE (8): R. M., *Essays on the Moral Concepts*, London, Macmillan, 1972.

HARE (9): R. M., *Applications of Moral Philosophy*, London, Macmillan, 1972.

HARMAN: G., *The Nature of Morality*, Oxford University Press, 1977.

HOFSTADTER AND DENNETT: D. R. Hofstadter and D. C. Dennett, eds., *The Mind's I*, Brighton, Harvester Press, 1981.

HOLLIS (1): M., 'Rational Man and Social Science', in R. Harrison, ed., *Rational Action: Studies in Philosophy and Social Science*, Cambridge University Press, 1979.

HOLLIS (2): M., *Models of Man*, Cambridge University Press, 1977.

HUME (1): D., *A Treatise of Human Nature*, Oxford, Clarendon Press, 1978.

HUME (2): D., 'The Sceptic', in Hume's *Essays*, Oxford University Press, 1963

HUME (3): D., *An Enquiry Concerning the Principles of Morals*, Oxford, Clarendon Press, 1975.

HURKA (1): T. M., 'Average Utilitarianisms', *Analysis* 42, No. 2, Mar. 1982.

HURKA (2): T. M., 'More Average Utilitarianisms', *Analysis* 42, June 1982

HURKA (3): T. M., 'Value and Population Size', *Ethics* 93, No. 3, Apr. 1983.

KANT: I., *Critique of Pure Reason*, trans. by N. Kemp Smith, London, Macmillan, 1964.

KAVKA (1): G., 'Rawls on Average and Total Utility', *Philosophical Studies* 27, Apr. 1975.

KAVKA (2): G., 'Some Paradoxes of Deterrence', *The Journal of Philosophy* 75, No. 6, June 1978.

KAVKA (3): G., 'Deterrence, Utility, and Rational Choice', *Theory and Decision* 12, 1980.

KAVKA (4): G., 'The Paradox of Future Individuals', *Philosophy and Public Affairs* 11, No. 2, Spring 1982.

KRIPKE: S. A., 'Naming and Necessity', in G. Harman and D. Davidson, eds., *Semantics of Natural Language*, Dordrecht, Reidel, 1972.

LESLIE: J., *Value and Existence*, Oxford, Basil Blackwell, 1979.

LEWIS (1): C. I., *An Analysis of Knowledge and Valuation*, La Salle, Ill., Open Court, 1946.

LEWIS (2): D. K., *Convention: A Philosophical Study*, Cambridge, Mass., Harvard University Press, 1969.

LEWIS (3): D. K., 'Survival and Identity', in RORTY .

LEWIS (4): H. D., *The Elusive Mind*, London, George Allen and Unwin, 1969.
LEWIS (5): H. D., *The Self and Immortality*, New York, Seabury Press, 1973.
LEWIS (6): H. D., *The Elusive Self*, London, Macmillan, 1982.
LICHTENBERG: G. C., *Schriften und Briefe*, Sudelbucher II, Carl Hanser Verlag, 1971.
LOCKE: J., *Essay Concerning Human Understanding*, partly reprinted in PERRY (1).
LYONS (1): D., *Forms and Limits of Utilitarianism*, Oxford, Clarendon Press, 1965.
LYONS (2): D., 'Human Rights and the General Welfare', *Philosophy and Public Affairs* 6, No. 2, Winter 1979.
MCDERMOTT: M., 'Utility and Population', *Philosophical Studies* 42, 1982.
MACKAYE: J., *The Economy of Happiness*, Boston, Little, Brown, 1906.
MACKIE (1): J. L., 'Sidgwick's Pessimism', *Philosophical Quarterly*, 1976,
MACKIE (2): J. L., *Ethics*, Harmondsworth, Penguin Books, 1977.
MACKIE (3): J. L., 'Rules and Norms', in Vol. 1 of *Selected Papers*, Oxford, Clarendon Press, forthcoming.
MACKIE (4): J. L., *Problems from Locke*, Oxford, Clarendon Press, 1976.
MACKIE (5): J. L., 'The Transcendental "I"', in VAN STRAATEN.
MACLEAN AND BROWN: D. MacLean and P. G. Brown, eds., *Energy and the Future*, Totowa, N.J., Rowman and Littlefield, 1983.
MCMAHAN (1): J. A., 'Problems of Population Theory', *Ethics* 92, No. 1, Oct. 1981.
MCMAHAN (2): J. A., 'Prudential Analogues', unpublished paper.
MCMAHAN (3): J. A., 'Nuclear Deterrence and Future Generations', in S. Lee and A. Cohen, eds., *Nuclear Weapons and the Future of Humanity*, Paterson, N.J., Littlefield Adams, forthcoming.
MADELL: G., *The Identity of the Self*, Edinburgh University Press, 1981.
MARGLIN: S., 'The Social Rate of Discount and the Optimal Rate of Investment', *Quarterly Journal of Economics*, 1963.
MARTIN AND DEUTSCHER: 'Remembering', *Philosophical Review*, 1966.
MEEHL: P., 'The Paradox of the Throw-Away Vote', *American Political Science Review* 71, 1977.
MILLER AND SARTORIUS: F. Miller and R. Sartorius, 'Population Policy and Public Goods', *Philosophy and Public Affairs* 8, No. 2, Winter 1979.
MILLER AND WILLIAMS: H. B. Miller and W. H. Williams, eds., *The Limits of Utilitarianism*, Minneapolis, University of Minnesota Press, 1982.
MONTEFIORE: A., ed., *Philosophy and Personal Relations*, London, Routledge and Kegan Paul, 1973.
MOORE: G. E., *Principia Ethica*, Cambridge University Press, 1903.
NABOKOV: V., *Glory*, London, Weidenfeld and Nicholson, 1971.
NAGEL (1): T., *The Possibility of Altruism*, Oxford University Press, 1970.
NAGEL (2): T., 'War and Massacre', *Philosophy and Public Affairs* 1, No. 2, Winter 1972.
NAGEL (3): T., 'The Limits of Objectivity', in *The Tanner Lectures on Human Values*, University of Utah Press, 1980.
NAGEL (4): T., *Mortal Questions*, Cambridge University Press, 1979.
NAGEL (5): T., 'Brain Bisection and the Unity of Consciousness', *Synthese* 22, 1971, reprinted in NAGEL (4) and in PERRY (1).
NARVESON (1): J., 'Utilitarianism and New Generations', *Mind* 76, Jan. 1967.
NARVESON (2): J., 'Moral Problems of Population', *The Monist*, 57, No. 1, Jan. 1973.
NEWMAN: Cardinal Newman, *Certain Difficulties Felt By Anglicans in Catholic Teaching*, London, 1885.
NIETZSCHE: F., *The Portable Nietzsche*, ed. and trans. by W. Kaufman, New York, Viking Press, 1963.
NORMAN: R., *Reasons for Actions*, Oxford, Basil Blackwell, 1971.
NOZICK (1): R., 'On Austrian Methodology'.
NOZICK (2): R., *Anarchy, State, and Utopia*, Oxford, Basil Blackwell, 1974.

NOZICK (3): R., *Philosophical Explanations*, Cambridge, Mass., Harvard University Press, 1981.

OLSON: M. Olson Jun., *The Logic of Collective Action*, Cambridge, Mass., Harvard University Press, 1965.

PARFIT (1): D., 'Personal Identity', *Philosophical Review* 80, No. 1, Jan. 1971, reprinted in PERRY (1).

PARFIT (2): D., 'On 'The Importance of Self-identity'', *The Journal of Philosophy* 68, No. 20, 21 Oct. 1971.

PARFIT (3): D., 'Later Selves and Moral Principles', in MONTEFIORE .

PARFIT (4): D., 'Innumerate Ethics', *Philosophy and Public Affairs* 7, No. 4, Summer 1978.

PARFIT (5): D, 'Prudence, Morality, and the Prisoner's Dilemma', *Proceedings of the British Academy* 65, 1979.

PARFIT (6): D., 'Personal Identity and Rationality', *Synthese* 53, 1982.

PARFIT (7): D., 'Future Generations: Further Problems', *Philosophy and Public Affairs* 11, No. 2, Spring 1982.

PARTRIDGE: E., ed., *Responsibilities to Future Generations*, Buffalo, N.Y., Prometheus Books, 1981.

PEACOCKE (1): C., 'Are Vague Predicates Incoherent?', *Synthese* 46, 1981.

PEACOCKE (2): C., *Sense and Content*, Oxford, Clarendon Press, 1983.

PEARS (1): D. F., 'Time, Truth, and Inference', in A. Flew, ed., *Essays in Conceptual Analysis*, London, Macmillan, 1966.

PEARS (2): D. F., 'Critical Study of *Individuals*, by P. F. Strawson', *Philosophical Quarterly* 11, 1961.

PENELHUM: T., 'The Importance of Self-identity', *The Journal of Philosophy* 68, No. 20, 21 Oct. 1971.

PERRY (1): J., ed., *Personal Identity*, Berkeley, University of California Press, 1975.

PERRY (2): J., *A Dialogue on Personal Identity and Immortality*, Indianapolis, Hackett, 1978.

PERRY (3): R. B., *General Theory of Value*, Cambridge, Mass., Harvard University Press, 1950.

PLATO (1): *Protagoras*.

PLATO (2): *Philebus*.

PROUST (1): M., *The Sweet Cheat Gone*, 1, trans. by C. K. Scott-Moncrieff, London, Chatto and Windus, 1949.

PROUST (2): M., *Within a Budding Grove*, trans. by C. K. Scott-Moncrieff, London, Chatto and Windus, 1949.

QUINE (1): W. V., reviewing Milton K. Munitz., ed., *Identity and Individuation*, in *The Journal of Philosophy*, 1972.

QUINE (2): W. V., 'What Price Bivalence?', *The Journal of Philosophy*, 1981.

QUINTON: A., 'The Soul', *The Journal of Philosophy*, 59, No. 15, July 1962, reprinted in PERRY (1).

RAHULA: W., *What the Buddha Taught*, New York, Grove Press, 1974.

RAILTON: P., 'Alienation and the Demands of Morality', *Philosophy and Public Affairs*, forthcoming.

RAPOPORT: A., *Fights, Games, and Debates*, Ann Arbor, University of Michigan Press, 1960.

RASHDALL: H., *The Theory of Good and Evil*, Oxford University Press, 1907.

RAVERAT: G., *Period Piece*, London, Faber and Faber, 1952.

RAWLS: J., *A Theory of Justice*, Cambridge, Mass., Harvard University Press, 1971.

RAY: C., 'Can We Travel Faster Than Light?', *Analysis* 42, Jan. 1982.

RAZ: J., *Practical Reason and Norms*, London, Hutchinson, 1975.

REGAN: D., *Utilitarianism and Co-operation*, Oxford, Clarendon Press, 1980.

REID: T., *Essays on the Intellectual Powers of Man*, first published in 1785, 'Of

Memory', chap. 4, reprinted in PERRY (1).
RESCHER: N., *Unselfishness*, University of Pittsburgh Press, 1975
RICHARDS: D. A. J., *A Theory of Reasons for Action*, Oxford, Clarendon Press, 1971.
RIKER AND ORDESHOOK: W. Riker and P. Ordeshook, 'A Theory of the Calculus of Voting', *American Political Science Review* 62, Mar. 1978.
ROBERTSON: Sir D., *Lectures on Economic Principles*, London, Collins, 1969.
RORTY: A., ed., *The Identities of Persons*, Berkeley, University of California Press, 1976.
ROSS (1): W. D., *The Right and the Good*, Oxford, The Clarendon Press, 1930.
ROSS (2): W. D., *The Foundations of Ethics*, Oxford, The Clarendon Press, 1939.
RUSSELL: B., 'On the Nature of Acquaintance', reprinted in R. C. Marsh, ed., *Logic and Knowledge*, London, Allen and Unwin, 1956.
SAINSBURY: R. M., 'In Defence of Degrees of Truth', unpublished paper.
SALMON: N. U., *Reference and Essence*, Oxford, Basil Blackwell, 1982.
SAMUELSON: P. A., *Economics*, New York, McGraw-Hill, 1970.
SCANLON (1): T. M., 'Preference and Urgency', *The Journal of Philosophy* 72, No. 19, 6 Nov. 1975.
SCANLON (2): T. M., 'Rights, Goals, and Fairness', in S. Hampshire, ed., *Public and Private Morality*, Cambridge University Press, 1978.
SCANLON (3): T. M., 'Contractualism and Utilitarianism', in SEN AND WILLIAMS .
SCHEFFLER (1): S., 'Moral Independence and the Original Position', *Philosophical Studies* 35, 1979.
SCHEFFLER (2): S., *The Rejection of Consequentialism*, Oxford, Clarendon Press, 1982.
SCHEFFLER (3): S., 'Ethics, Personal Identity, and Ideals of the Person', *Canadian Journal of Philosophy* 12, No. 2, June 1982.
SCHELL: J., *The Fate of the Earth*, New York, Avon Books, 1982.
SCHELLING (1): T., *The Strategy of Conflict*, Cambridge, Mass., Harvard University Press, 1960.
SCHELLING (2): T., 'Hockey Helmets, Concealed Weapons, and Daylight Saving', *The Journal of Conflict Resolution*, Sept. 1973.
SCHNEEWIND: J. B., *Sidgwick's Ethics and Victorian Moral Philosophy*, Oxford University Press, 1977.
SCHUELER: G. F., 'Nagel on the Rationality of Prudence', *Philosophical Studies* 29, 1976.
SEN (1): A. K., *Collective Choice and Social Welfare*, San Francisco, Holden Day, 1970.
SEN (2): A. K., *Behaviour and the Concept of Preference*, London School of Economics, 1973.
SEN (3): A. K., 'Isolation, Assurance, and the Social Rate of Discount', *Quarterly Journal of Economics* 81, 1967.
SEN (4): A. K., 'Rational Fools: A Critique of the Behavioral Foundations of Economic Theory', *Philosophy and Public Affairs* 6, No. 4, Summer 1977.
SEN (5): A. K., 'Utilitarianism and Welfarism', *The Journal of Philosophy* 76, No. 9, Sept. 1979.
SEN (6): A. K., *On Economic Inequality*, Oxford, Clarendon Press, 1973.
SEN (7): A. K., 'Choice, Orderings, and Morality', in S. Korner, ed., *Practical Reason*, Oxford University Press, 1974.
SEN AND RUNCIMAN: A. K. Sen and W. G. Runciman, 'Games, Justice and the General Will', *Mind* 74, Oct. 1965.
SEN AND WILLIAMS: A. K. Sen and B. Williams, *Utilitarianism and Beyond*, Cambridge University Press, 1982.
SHOEMAKER (1): S., *Self-Knowledge and Self-Identity*, Ithaca, N.Y., Cornell University Press, 1963.
SHOEMAKER (2): S., 'Persons and Their Pasts', *American Philosophical Quarterly* 7, 1970.

SHOEMAKER (3): S., 'Wiggins on Identity', *Philosophical Review* 79, 1970.
SHORTER: J. M., 'More About Bodily Continuity and Personal Identity', *Analysis* 22, 1961-2.
SIDGWICK (1): H., *The Methods of Ethics*, London, Macmillan, 1907.
SIDGWICK (2): A. S. and E. M. S. Sidgwick, *Henry Sidgwick, a Memoir*, London, Macmillan, 1906.
SIKORA AND BARRY: R. I. Sikora and B. Barry, eds., *Obligations to Future Generations*, Philadelphia, Temple University Press, 1978.
SINGER (1): M., 'The Many Methods of Sidgwick's Ethics', *The Monist* 58, 1974.
SINGER (2): P., *The Expanding Circle*, New York, Farrar, Straus, and Giroux, 1981.
SMART (1): B., 'Diachronous and Synchronous Selves', *Canadian Journal of Philosophy* 6, No. 1, Mar. 1976.
SMART (2): J. J. C., ed., *Problems of Space and Time*, New York, Macmillan, 1976.
SMART AND WILLIAMS: J. J. C. Smart and B. Williams, *Utilitarianism: For and Against*, Cambridge University Press, 1973.
SOBEL (1): J. H., 'Interaction Problems for Utility Maximizers', *Canadian Journal of Philosophy* 4, No. 4, June 1975.
SOBEL (2): J. H., 'The Need for Coercion', in J. Pennock and H. Chapman, eds., *Coercion*, Chicago, Nomos, 1972.
SOLZHENITSYN: A., *The First Circle*, New York, Bantam Books, 1969.
SPERRY: R. W., in J. C. Eccles, ed., *Brain and Conscious Experience*, Berlin, Springer Verlag, 1966.
STRANG: C., 'What if Everyone Did That?', in THOMSON AND DWORKIN.
STRAWSON (1): P. F., *Individuals*, London, Methuen, 1959.
STRAWSON (2): P. F., *The Bounds of Sense*, London, Methuen, 1966.
STROTZ: 'Inconsistency and Myopia in Dynamic Utility Maximization', *Review of Economic Studies*, 1955-6.
SUMNER (1): L. W., 'The Good and the Right', *Canadian Journal of Philosophy*, Suppl. Vol. on John Stuart Mill and Utilitarianism, 1979.
SUMNER (2): L. W., *Abortion and Moral Theory*, Princeton University Press, 1981.
SWINBURNE: R. G., 'Personal Identity', *Proceedings of the Aristotelian Society* 74, 1973-4.
TEMKIN: L., 'Inequality', Ph.D. dissertation, Princeton University, 1982.
THOMSON AND DWORKIN: J. J. Thomson and G. Dworkin, eds., *Ethics*, New York, Harper and Row, 1968.
TOOLEY: M., *Abortion and Infanticide*, Oxford, Clarendon Press, 1983.
TREBLICOT: J., 'Aprudentialism', *American Philosophical Quarterly* 11, No. 3, July 1974.
ULLMAN-MARGALIT: E., *The Emergence of Norms*, Oxford, Clarendon Press, 1977.
UNGER (1): P., 'Why There Are No People', *Midwest Studies in Philosophy* 4, 1979.
UNGER (2): P., 'I Do Not Exist', in G. F. MacDonald, ed., *Perception and Identity*, Ithaca, N.Y., Cornell University Press, 1979.
VAN STRAATEN: Z., *Philosophical Subjects, Essays presented to P. F. Strawson*, Oxford, Clarendon Press, 1980.
WACHSBERG: M., *Personal Identity, the Nature of Persons, and Ethical Theory*, Ph.D. Dissertation, Princeton University, 1983.
WARNOCK: G. J., *The Object of Morality*, London, Methuen, 1971.
WATKIN: J., 'Imperfect Rationality', in R. Borger and F. Cioffi, eds., *Explanation in the Behavioural Sciences*, Cambridge University Press, 1970.
WIGGINS (1): D., *Identity and Spatio-Temporal Continuity*, Oxford, Basil Blackwell, 1967.
WIGGINS (2): D., 'Essentialism, Continuity, and Identity', *Synthese* 23, 1974.
WIGGINS (3): D., *Sameness and Substance*, Oxford, Basil Blackwell, 1980.
WIGGINS (4): D., 'Locke, Butler and the Stream of Consciousness', in RORTY.

WIGGINS (5): D., 'Identity, Designation, Essentialism and Physicalism', *Philosophia* 5, Nos. 1-2, Jan.-Apr. 1975.

WIGGINS (6): D., 'The Concern to Survive', *Midwest Studies in Philosophy* 4, 1979.

WILLIAMS (1): B., 'A Critique of Utilitarianism', in SMART AND WILLIAMS.

WILLIAMS (2): B., *Problems of the Self*, Cambridge University Press, 1973.

WILLIAMS (3): B., 'Persons, Character, and Morality', in RORTY.

WILLIAMS (4): B., *Descartes*, Harmondsworth, Penguin Books, 1978.

WILLIAMS (5): B., 'Internal and External Reasons', in R. Harrison, ed., *Rational Action: Studies in Philosophy and Social Science*, Cambridge University Press, 1979.

WILLIAMS (6): B., *Moral Luck*, Cambridge University Press, 1981.

WILLIAMS (7): B., 'The Point of View of the Universe: Sidgwick and the Ambitions of Ethics', *The Cambridge Review* 7 May 1982.

WILLIAMS (8): B., 'The Self and the Future', *Philosophical Review* 79, No. 2, Apr. 1970, reprinted in WILLIAMS (2).

WILLIAMS (9): B., 'Bodily Continuity and Personal Identity', *Analysis* 20, No. 5, reprinted in WILLIAMS (2).

WOODS: M., 'Reference and Self-Identification', *The Journal of Philosophy*, 1968.

译后记

2004年《理与人》中文版由上海译文出版社出版，迄今已经十年有余。在这十多年的时间内，特别是在国外，《理与人》在哲学、伦理学和政治学等领域产生了巨大的影响。在这十多年的时间内，帕菲特对于他在《理与人》中所表述的相关理论的各种反馈，对于其中的某些思想也有着进一步的调适和细化。是故，在《理与人》中文版重印之际，本人感到在"译后记"中有必要再多说几句。

德里克·帕菲特（Derek Parfit，1942— ）是英国当代著名道德哲学家、心灵哲学家，专长个人同一性、合理性等伦理学问题。相较于麦金泰尔等人，国内学界对于德里克·帕菲特的著作和思想的译介与研究较晚，但事实上他比当代任何西方道德哲学家或伦理学家与中国都更加有缘分：1942年帕菲特出生在中国成都。帕菲特的父母诺曼·帕菲特和杰茜·帕菲特当年都是在中国西部的教会医院传授预防医学的医生。帕菲特出生后的次年，一家人回国定居剑桥。学龄时期帕菲特被送往伊顿公学接受教育，后来于1965年获得牛津大学现代史硕士学位。帕菲特1965—1966年前往美国哥伦比亚大学和哈佛大学从事研究，期间放弃历史研究而转向哲学。

帕菲特的主要学术生涯是在牛津大学度过的。1984—2010年期间他担任牛津大学万灵学院高级研究员，现为牛津大学荣休高级研究员。荣休后，帕菲特还兼任美国纽约大学"全球杰出哲学教授"讲席永久客座教授，以及哈佛大学、普林斯顿大学、天普大学和莱斯大学等著名大学的客座教授。作为当代伦理学的领军人物之一，帕菲特还是赫赫有名的英国不列颠学会会员（the Fellow of British Academy），

以及美国艺术与科学研究院会员（Fellow of the American Academy of Arts and Science）。

《理与人》（1984，1986，1987）是帕菲特的代表作，被西方学界公认为过去几十年间最有影响的伦理学和哲学著作之一。《理与人》共二十章，分为四编，另有附录。正如书名所示，该书主要围绕理（行动的理由）与人（行动者）两个主题展开。首先，在第一编中处理了在他看来以自利论为代表的传统道德理论是“自败的理论”，在与自利论彼此交锋的关系中论述了道德论和当前目标论等理论；其次，鉴于自利论是否合理的问题与时间中展开的任何理由的力量相关，在第二编中处理了“合理性与时间”问题，并通过大量例证否证了历史上大多数西方人所持的自利论，复鉴于合理性和道德问题总与作为行为主体的人相关，在第三编中展开了关于“个人同一性”的讨论，在第四编中展开了“未来的世世代代”的讨论。通过以上四编，《理与人》以强有力的论证动摇了西方人传统上在“合理性”、“道德”和“个人同一性”等问题上的某些根深蒂固的信念和立场。

帕菲特认为，大多数人应该改变传统的有关个人本质和历时的个人同一性的观点；鉴于自利论和后果论经常是间接自败的，所以应当拒斥合理性的自利论；不仅如此，绝大多数人持有的道德观念（常识道德）是经常直接地、集体性地自败的，而且在行动如何影响他人，特别是行动影响未来的人等方面人们犯了严重错误，因而应当改变我们目前的许多行为方式，并且尝试建立一种全新的、并非基于宗教的非宗教伦理学。

在英美国家，伦理和道德方面的教学与研究崇尚分析与推理，《理与人》就是一部典型的道德分析与推理著作。当代著名哲学家斯特劳森在《纽约书评》中评价说：“此类著作中鲜有在范围、丰富程度、想象力和推理的说服力方面可与之相媲美者。”另一位著名哲学家舍弗勒在《时代周刊》中撰文指出：“《理与人》可能是自百年前西季威克发表《伦理学方法》以来功利主义传统中有实质性的道德哲学的

最伟大著作。”该书不仅荣登美国《哲学论坛》1999年12月发表的20世纪哲学经典的调查报告的25本经典之一，而且是英美大学中理论伦理学和应用伦理学的必读图书之一。事实上，《理与人》一直占据各种有关最有影响的伦理学著作的报告和排行的前列。

然而，本书与中国学界的学术偏好和氛围一度似有暂时不合之处。这主要是因为该书读来有几大困难：一是此书作为从分析哲学出发来研究功利主义传统的巨著，既广征博引，又条分缕析，分析哲学的方法让人生畏；二是用无数想象独到的事例对于人们所秉持的有关合理性、道德和个人同一性等传统信念提出挑战，要求对作为靶子的相关理论有一定的在先知识；三是作者在有关论点的阐发上并非直抒己见，而是在不同的观点之间借力打力，要求有相当强的逻辑推理能力；四是该书行文和逻辑发展并非像坡道一样平滑，而是像楼梯那样具有阶梯性，要求读者在阅读中自我补足命题之间某种程度的跳跃环节。

除了《理与人》之外，帕菲特还有不少代表性论文。其中最为重要的论文有“平等或优先？”（1995）、“理由和动机”（1997）、“为何任何？为何唯此？”（1998）、“经验、主体和概念框架”（1999）、“规范性”（2006）。另外，2011年牛津大学出版社推出了帕菲特的两卷本近作《何事体大》（On What Matters），其中的第一卷已有中文译本（《论重要之事》）。据悉，2017年牛津大学出版社将推出《何事体大》的第三卷。鉴于《何事体大》中的问题意识和相关论题源于《理与人》及其后续讨论，实际上《理与人》已成为人们读懂《何事体大》的必要前提。

《何事体大》作为学界期待已久的《理与人》的姊妹篇，通过对康德伦理学、契约主义和后果论三大系统道德论的批判性稽考，结合《理与人》出版后读者的反应和学者的诘难，对理由与合理性等论题进行了反馈式全新阐释，与《理与人》一道为其赢得了绝大多数哲学家眼中“最伟大的在世道德哲学家”的地位。“因其在关乎个人同一性、

未来的世世代代和道德理论结构方面的突破性贡献”，帕菲特还在2014年荣获每三年一届的罗夫·肖克奖（Rolf Schock Prizes）所设的“逻辑与哲学”、“数学”、“视觉艺术”和“音乐艺术”四项奖项中的“逻辑与哲学奖”。

《理与人》的出版和再版，凝结着上海译文出版社有关领导和编审人员的辛勤劳动和智慧，在此表示衷心的感谢。此外，我还要特别鸣谢《理与人》中译本的校对者、我的博士导师黄颂杰教授，译者怀着弟子的敬意，谨以再版的这部译作献给即将迎来八十寿诞的黄颂杰教授。

我始终认为，在某种意义上，翻译无非是把原作者的真知灼见一字一句地再现给读者看，但同时也不免把翻译者的孤陋寡闻一点一滴地露给读者看。尽管借此重印之际，译者校阅了译文，并且就读者商榷的两处译文进行了修订，但是仍然恳请伦理学界和翻译学界的前辈、同侪、青年才俊和广大读者对于仍有可能存在的不足提出宝贵意见。

王新生

2016年11月21日于复旦大学光华楼